SECOND WORKSHOP ON RAISING DATA USING THE RDBES AND TAF (WKRDBES RAISE&TAF2)

VOLUME 5 | ISSUE 109

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM

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

ISSN number: 2618-1371

ICES Scientific Reports

Volume 5 | Issue 109

SECOND WORKSHOP ON RAISING DATA USING THE RDBES AND TAF (WKRDBES RAISE&TAF2)

Recommended format for purpose of citation:

ICES. 2023. Second Workshop on Raising Data using the RDBES and TAF (WKRDBES Raise&TAF2). ICES Scientific Reports. 5:109. 71 pp. https://doi.org/10.17895/ices.pub.24648546

Editors

David Currie • Edvin Fuglebakk

Authors

Maciej Adamowicz • David Currie • Afra Egan • Edvin Fuglebakk • Ana Cláudia Fernandes • Richard Meitern • Nuno Prista • Eros Quesada • Yves Reecht • Lies Vansteenbrugge • Maria Ching Villanueva

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The Second Workshop on Raising Data using the RDBES and TAF (WKRDBES Raise&TAF2) met to reproduce estimates of commercial catch data using the Regional Database and Estimation System (RDBES) and the Transparent Assessment Framework (TAF). Both the direct input to stock assessments (stock coordination) and the upstream national estimates were attempted to be reproduced for several stocks. The workshop provided examples of successful reproductions, continuing the reproduction efforts started at WKRDBES_Raise&TAF which convened in 2022. Both kinds of estimates were also implemented in TAF, demonstrating the feasibility of RDBES/TAF to facilitate transparent computation of accepted estimation practices. The workshop also identified issues that would prevent the RDBES/TAF approach from being acceptable to many participating institutions due to data confidentiality concerns and unclear data quality declarations and proposed possible solutions.

ii Expert group information

Expert group name	Second Workshop on Raising Data using the RDBES and TAF (WKRDBES Raise&TAF2)
Expert group cycle	Annual
Year cycle started	2023
Reporting year in cycle	1/1
Chairs	David Currie, Ireland
	Edvin Fuglebakk, Norway
Meeting venue and dates	2 – 6 October 2023, Online (45 participants)

1 Introduction

The main aim of the WKRDBES Raise&TAF2 workshop is to test if new workflows for commercial catch data used in ICES advice can reproduce the results of previous workflows. Specifically, the workflows implemented with the transparent assessment framework (TAF) and the Regional Database and Estimation System (RDBES) are compared with the output of corresponding workflows that institutions have submitted to InterCatch (national estimates), or that has been implemented in InterCatch (stock estimates).

This workshop builds upon the previous WKRDBES-Raise&TAF and WKRDBES-Raise&TAF_Flow workshops and its focus is on re-producing current practices as closely as possible within RDBES/TAF rather than developing new estimation methodologies.

The Regional Database and Estimation System (RDBES) is being developed with the ambition to replace current databases supporting archiving of commercial fisheries data and produce stock assessment input, that is the 'Regional Database' (RDB) and 'InterCatch'. The governance group for the RDBES development (WGRDBESGOV) anticipates that the new system will be developed until 2024, at which point it will be ready to replace RDB and InterCatch. An important prerequisite for phasing out RDB and InterCatch is to demonstrate that RDBES can provide sufficient support for current estimation protocols. This is well demonstrated if RDBES and TAF can be used to reproduce the output from current protocols.

This workshop has the following terms of reference:

a) National estimation using RDBES and TAF

> i) Reproduce the 2023 upload (2022 data) to InterCatch by producing R-scripts that raise national data extracted from the RDBES format to national level estimates. Compare with previously uploaded estimates; (This ToR is a continuation of ToR a from WKRDBES-RAISE_TAF 2022.)
>
> ii) Set up national TAF repositories and produce R-scripts for generic, standard approaches; The work should build on the outcome of WKRaise&TAF_Flow.

b) Stock coordination using RDBES and TAF

> i) Reproduce the 2023 stock coordination (2022 data) previously done in InterCatch, with the R-scripts that run on ToR a output. Compare with previously achieved estimates. (This ToR is a continuation of ToR b from WKRDBES-Raise&TAF 2022.)
>
> ii) Set up stock estimation TAF repositories and produce R-scripts for generic, standard approaches; The work should build on the outcome of WKRaise&TAF_Flow.

1.1 Background

The current system for submitting fisheries dependent data to stock assessment was described in WGRDBESGOV 2021 https://doi.org/10.17895/ices.pub.21133372.v2 and is reproduced below.

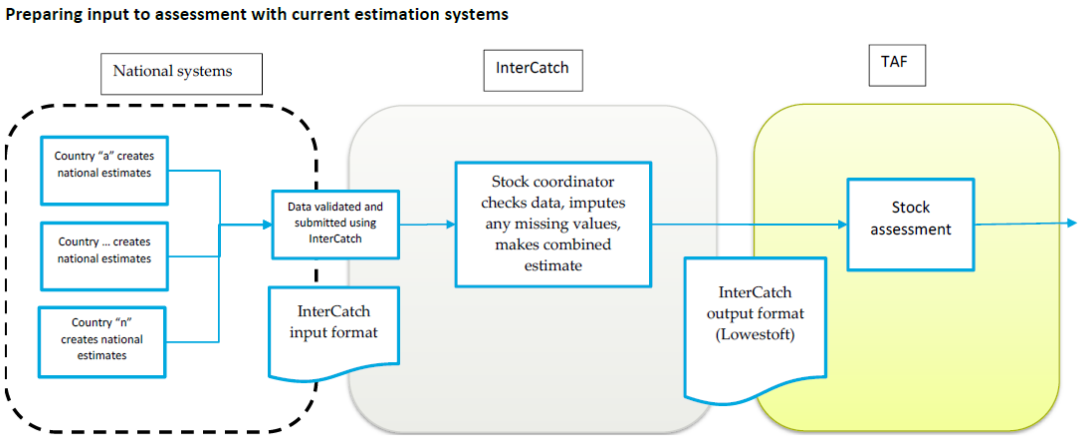


Figure 1 Current estimation systems.

This is contrasted with the proposed pathway using RDBES/TAF:

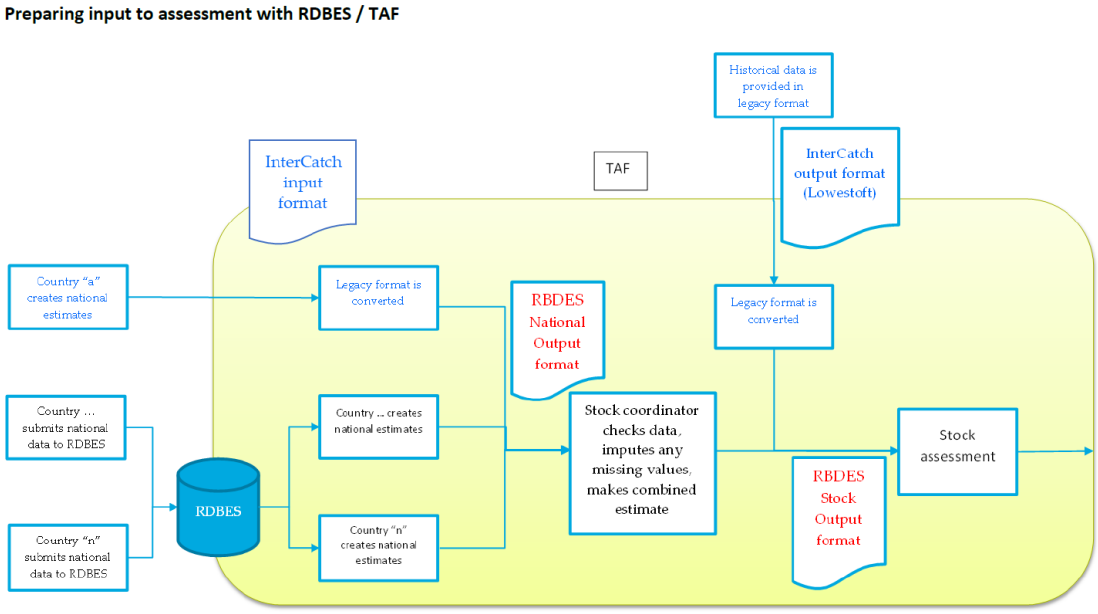


Figure 2 Proposed RDBES/TAF estimation.

A broad picture of how TAF repositories can be linked together was outlined by WGTAFGOV in their manifesto https://github.com/ices-eg/wg_WGTAFGOV

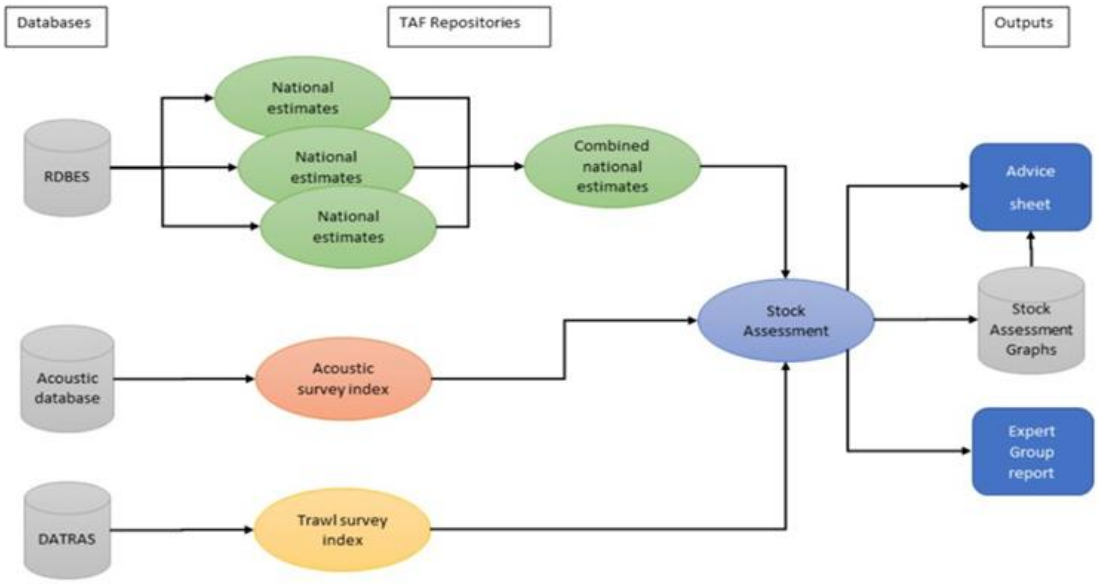


Figure 3 Linking TAF repositories.

The first WKRDBES-Raise&TAF workshop https://doi.org/10.17895/ices.pub.21995141.v1 gave a more detailed examination of this TAF architecture using the example of cod 27.21:

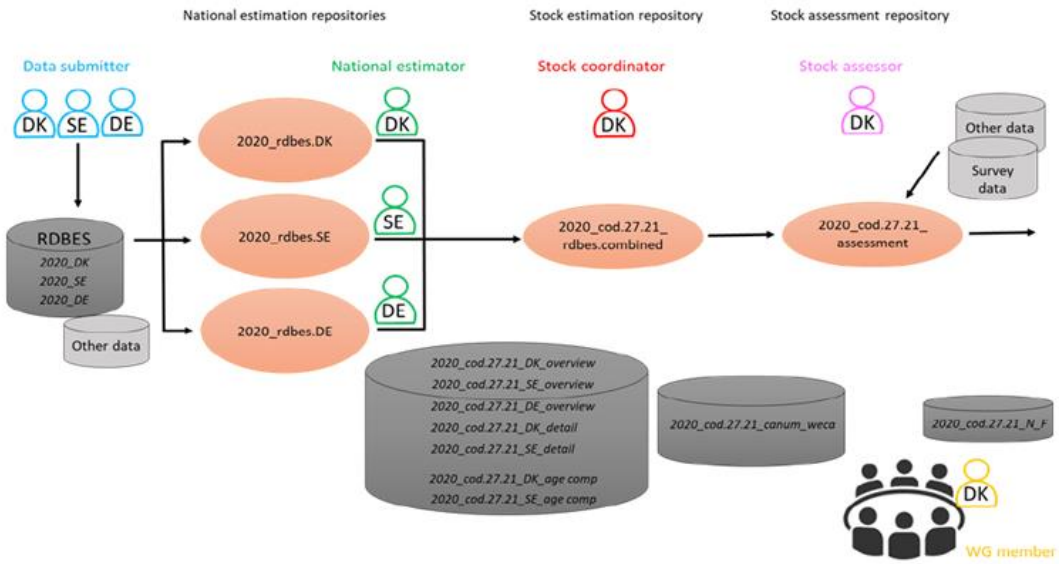


Figure 4 RDBES/TAF cod 27.21 example.

The RDBES is a structured, relational database but outputs within TAF generally exist as files within a repository. A TAF database exists which allows these outputs to be tracked and managed - each type of TAF output can be recorded within the database along with required metadata. It is then possible to request a particular output - the TAF database will decide whether you can have access to it based on your assigned roles, and the metadata in the TAF database.

2 Reproduction of national estimates

The progress towards reproducing national estimates is summarised below – this work is aimed at fulfilling ToR a.

Table 1 Progress on reproduction of national estimates

Participants	Prior effort	Stocks	Quantity	TAF progress	Implementation	Reproduction
Belgium	R&T 22	sol.27.7fg	LAN,DAN	started	Partial	Partial
Denmark	None	cod.27.21	LLN, LAN	Started/Issues	Partial/Issues	Partial
Estonia	None	SPR 27.3.d.28-32	LAN	Complete	Partial	Partial
Finland	None	spr.27.22-32 her.27.30-31 her.27.25-2932	LAN, LLN	Started	Started	Started
France	None	sol.27.8ab syc.27.3a47d	LLN	None	Partial	Partial
Germany	R&T 22	whb.27.1-91214 (datacall WGWIDE)	LLN, LAN	Complete	Partial	Partial, Issues
Ireland	R&T 22 days	Ple.27.7h-k, ple.27.fg, had.27.7b-k, whg.27.7a	LAN, LLN,	Started	Complete	Reproduced
Ireland	days	ple.27.7h-k	DAN, DLN, DB	None	Started	Started
Ireland	R&T 22 days	Hom.27.2a4a5b6a7a-ce-k8, mac.27.nea, whb.27.1-91214	LAN	None	Complete	Reproduced
Latvia	none	her.27.28	LAN	Started	Started	Started
Netherlands	none	mac.27.nea	LAN,LLN	None	Complete	Partial
Norway	R&T_22	pok.27.3a46	LAN	-	-	-
Poland	R&T_22 days	ple.27.24-32	LAN	Partial	Complete	Partial
Spain, AZTI	R&T_22 weeks	bss.27.8ab, bss.27.8c9a, hke.27.3a46-8abd, hke.27.8c9a, sol.27.8ab, mac.27.nea, hom.27.2a4a5b6a7a-ce-k8	*LLN, DLN*	Not started	*Partial*	Partial
Spain, IEO	R&T_22	hke.27.8c9a	LLN	Not started	Partial	Partial
Sweden	R&T_22	Most demersal stocks	DB	Not started	Partial/Issues	Partial/Issues
UK (England)	R&T_22		DB and DLN	Started	Started/Issues	Started
UK (Scotland)	R&T_22	mac.27.nea	LAN	Partial - issues	Complete	Issues

Prior effort: If any work was done in adapting the estimate(s) to RDBES and TAF before the workshop. Use categories: 'None', 'days', 'weeks''. The text "R&T_22" indicates that previous work was done during the Raise&TAF workshop in 2022.

Stocks: stocks reproduced, uses ICES stock codes

Quantity: Which quantity was estimated. *LAN*: landings-at-age in numbers: total landings in numbers by age groups

- *LLN*: landings-at-length in numbers: total landings in numbers by length groups
- *DAN*: discards-at-age in numbers: total discard in numbers by age groups
- *DLN*: discards-at-length in numbers: total discard in numbers by length groups
- DB: total discards in weight: total discards in weight, not decomposed by age or length

TAF progress: To what extent TAF was used in the reproduction.

- *Started*: Did not organise code according to TAF standards.
- *Partial*: the estimate was organised with the standard TAF directories and scripts (icesTAF::taf.skeleton)
- *Complete*: the estimate can be re-run with icesTAF::sourceAll

Implementation: To what extent the implementation of estimator was done:

- *Started:* the implementation did not progress far enough to decide if the RDBES data model provides the necessary information
- *Issues:* the implementation could not be completed because of issues with the RDBES data model.
- *Partial:* not all code necessary to run estimates was implemented, but sufficient to test the RDBES and no issues with the RDBES encountered.
- *Complete*: All code necessary to run estimates was implemented.

Reproduction: To what extent the reproduction was successful:

- *Started*: Did not progress far enough to conclude on reproduction.
- *Reproduced*: Results are reproduced to the satisfaction of the person implementing the reproduction.
- *Partial*: Results are not reproduced to the satisfaction of the person implementing the reproduction, but the reasons why are thought to not be limitations of the RDBES data model.
- *Issues*: Results are not reproduced to the satisfaction of the person implementing the reproduction, and it the reasons why are thought to be limitations of the RDBES data model.

2.1 Report from reproduction studies

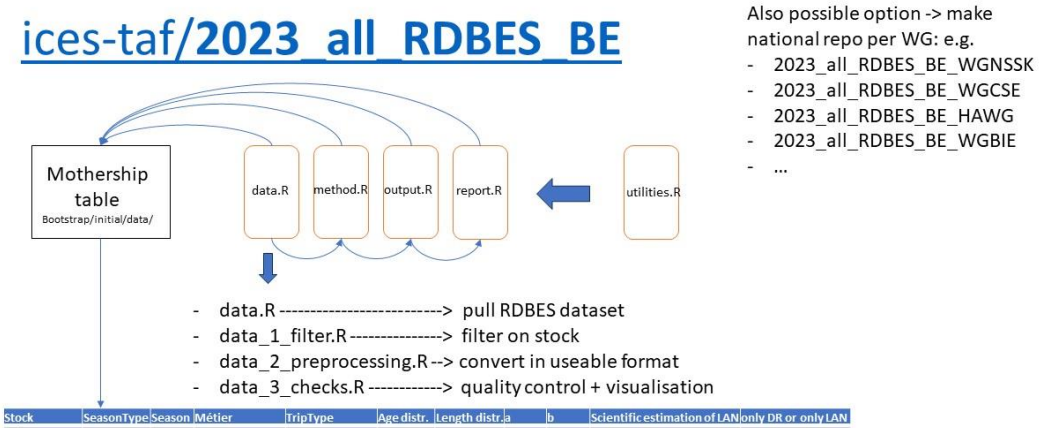
Belgium

Reproduction of national estimates: Belgium is rewriting the present estimation routines in R based on the RDBES format as input. Currently, the estimation is done in R using the r pack-age COST, now we are developing our own functions to do the raising procedure. We reconstructed discard raising, length weight and age length keys, raised length frequencies to numbers at length and age, discard and landings. We are still comparing the estimates using the currently used scripts for raising, and the newly developed scripts using the data in the RDBES format.

TAF progress: Belgium brainstormed about a way to structure its national TAF folder and presented this to the group. The format was however not implemented yet and remains a theoretical framework. Nevertheless, we hope to continue working on this in upcoming workshops.

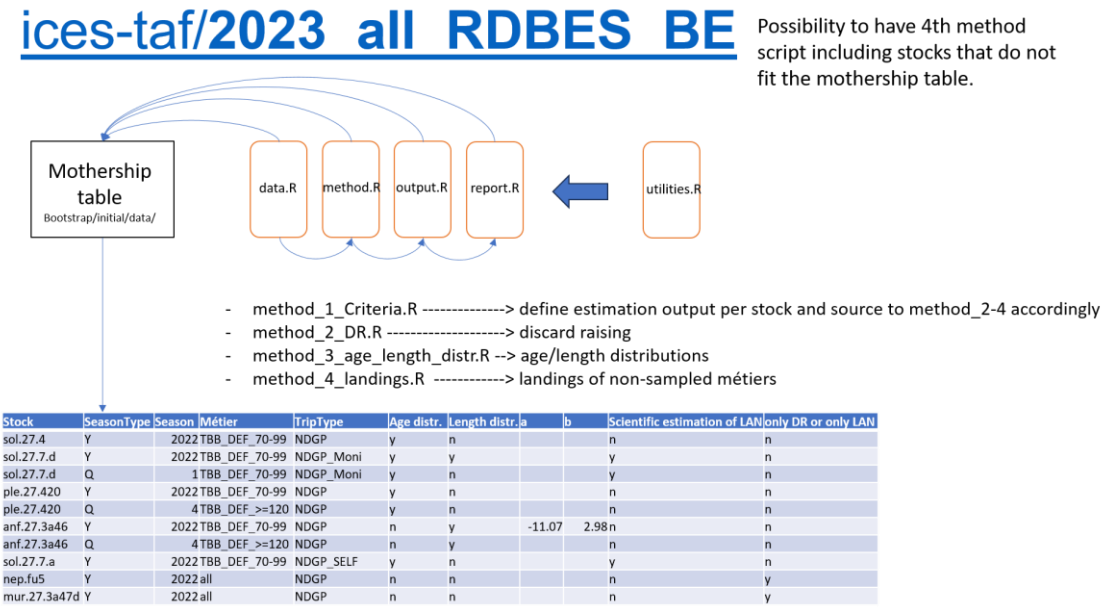
The general idea of this TAF format is to import a "mothership table" containing all stocks for which data is requested and their specifics to which the data.R, method.R, output.R and report.R would be able to call upon. Depending on the specifics, different R scripts will be sourced.

For the data part (see figure below), we plan to have 4 scripts: 1) data.R which will pull the data from the national RDBES database, 2) data_1_filter.R where the data will be filtered on each stock present in the mothership table, 3) data_2_preprocessing.R where the filtered data will be converted in a useable format for raising and 4) data_3_checks.R where a number of quality control and visualisation steps will be done.



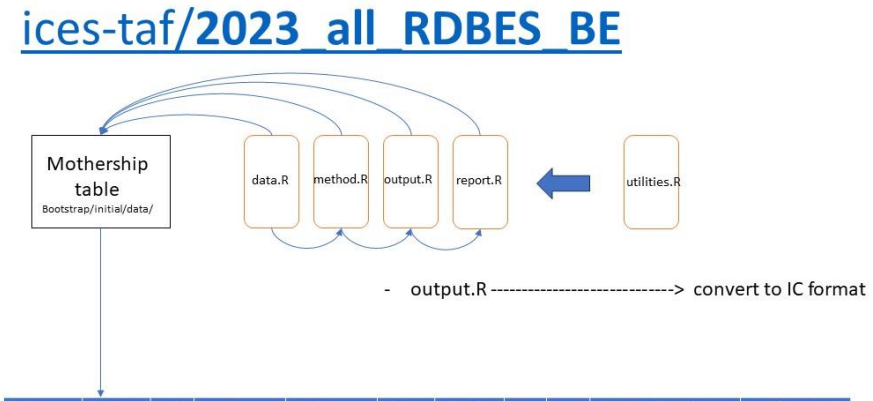
Stock	SeasonType	Season	Métier	TripType	Age distr.	Length distr.	a	b	Scientific estimation of LAN	only DR or only LAN
sol.27.4	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
sol.27.7.d	Y	2022	TBB_DEF_70-99	NDGP_Moni	y	y			y	n
sol.27.7.d	Q	1	TBB_DEF_70-99	NDGP_Moni	y	n			y	n
ple.27.420	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
ple.27.420	Q	4	TBB_DEF_>=120	NDGP	y	n			n	n
anf.27.3a46	Y	2022	TBB_DEF_70-99	NDGP	n	y	-11.07	2.98	n	n
anf.27.3a46	Q	4	TBB_DEF_>=120	NDGP	n	y			n	n
sol.27.7.a	Y	2022	TBB_DEF_70-99	NDGP_SELF	y	n			y	n
nep.fu5	Y	2022	all	NDGP	n	n			n	y
mur.27.3a47d	Y	2022	all	NDGP	n	n			n	y

For the method part (see figure below), we plan to have 4 scripts. The first one, method_1_Criteria.R is where the estimation output (discard quantity, landings quantity, age distribution and length distribution) will be assigned to each stock in the "mothership table". Based on the method defined and the specifics in that table one or more of the following method scripts will be sourced: 2) method_2_DR.R where a discard rate will be calculated for the stocks it applies, 3) method_3_age_length_distr.R where distributions of age or length will be calculated, 4) method_4_landings.R where landings of non-sampled métiers will be aggregated. Stocks with very specific and rare requirements could be directed to a 5th script, where the necessary information is being calculated.



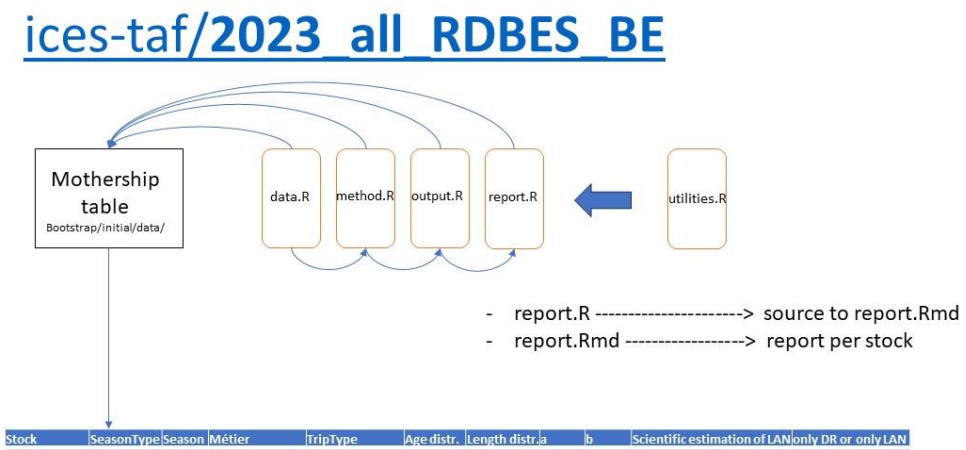
Stock	SeasonType	Season	Métier	TripType	Age distr.	Length distr.	a	b	Scientific estimation of LAN	only DR or only LAN
sol.27.4	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
sol.27.7.d	Y	2022	TBB_DEF_70-99	NDGP_Moni	y	y			y	n
sol.27.7.d	Q	1	TBB_DEF_70-99	NDGP_Moni	y	n			y	n
ple.27.420	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
ple.27.420	Q	4	TBB_DEF_>=120	NDGP	y	n			n	n
anf.27.3a46	Y	2022	TBB_DEF_70-99	NDGP	n	y	-11.07	2.98	n	n
anf.27.3a46	Q	4	TBB_DEF_>=120	NDGP	n	y			n	n
sol.27.7.a	Y	2022	TBB_DEF_70-99	NDGP_SELF	y	n			y	n
nep.fu5	Y	2022	all	NDGP	n	n			n	y
mur.27.3a47d	Y	2022	all	NDGP	n	n			n	y

For the output part (see figure below), we plan to have one script (output.R) that calls upon the output of the method scripts and converts the data to the currently used InterCatch (IC) format (HI, SI, SD).



Stock	SeasonType	Season	Métier	TripType	Age distr.	Length distr.	a	b	Scientific estimation of LAN	only DR or only LAN
sol.27.4	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
sol.27.7.d	Y	2022	TBB_DEF_70-99	NDGP_Moni	y	y			y	n
sol.27.7.d	Q	1	TBB_DEF_70-99	NDGP_Moni	y	n			y	n
ple.27.420	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
ple.27.420	Q	4	TBB_DEF_>=120	NDGP	y	n			n	n
anf.27.3a46	Y	2022	TBB_DEF_70-99	NDGP	n	y	-11.07	2.98	n	n
anf.27.3a46	Q	4	TBB_DEF_>=120	NDGP	n	y			n	n
sol.27.7.a	Y	2022	TBB_DEF_70-99	NDGP_SELF	y	n			y	n
nep.fu5	Y	2022	all	NDGP	n	n			n	y
mur.27.3a47d	Y	2022	all	NDGP	n	n			n	y

For the report part (see figure below), we will have 2 scripts: a report.R where we will source to a report.Rmd markdown file where a short summary report is generated for each stock.



Stock	SeasonType	Season	Métier	TripType	Age distr.	Length distr.	a	b	Scientific estimation of LAN	only DR or only LAN
sol.27.4	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
sol.27.7.d	Y	2022	TBB_DEF_70-99	NDGP_Moni	y	y			y	n
sol.27.7.d	Q	1	TBB_DEF_70-99	NDGP_Moni	y	n			y	n
ple.27.420	Y	2022	TBB_DEF_70-99	NDGP	y	n			n	n
ple.27.420	Q	4	TBB_DEF_>=120	NDGP	y	n			n	n
anf.27.3a46	Y	2022	TBB_DEF_70-99	NDGP	n	y	-11.07	2.98	n	n
anf.27.3a46	Q	4	TBB_DEF_>=120	NDGP	n	y			n	n
sol.27.7.a	Y	2022	TBB_DEF_70-99	NDGP_SELF	y	n			y	n
nep.fu5	Y	2022	all	NDGP	n	n			n	y
mur.27.3a47d	Y	2022	all	NDGP	n	n			n	y

Denmark

Reproduction of national estimates

Denmark started rewriting our present estimation routines in R based on the RDBES format as input. Presently the estimation is done in SAS with a format very similar to the RDB format as input. The focus at the WK was on Landings and even though the implementation is not fully done the preliminary result looks very promising and overall, the RDBES has the information needed, but we have an issue with the inclusion of the correction for overweight in boxes, see below.

Issue

The RDBES data model supports reporting of correction for overweight in boxes in the CLexplainDifference, but we find that the inclusion of this, in combination with all the other types of corrections we make on the way from official weight to scientific weight, makes it difficult to comprehend the reasoning behind the difference, e.g. at lot of the lines would be a combination of a lot of the codes present. We have created two issues relating to this at https://github.com/ices-tools-dev/RDBES/issues, #184 and #198

TAF

Preliminary ideas of how to set up of our national estimation repo was tested to see if the structure support our current way of working, see
https://github.com/ices-taf/2023_all_RDBES_DK/blob/main/national_repo_test_setup_1.html

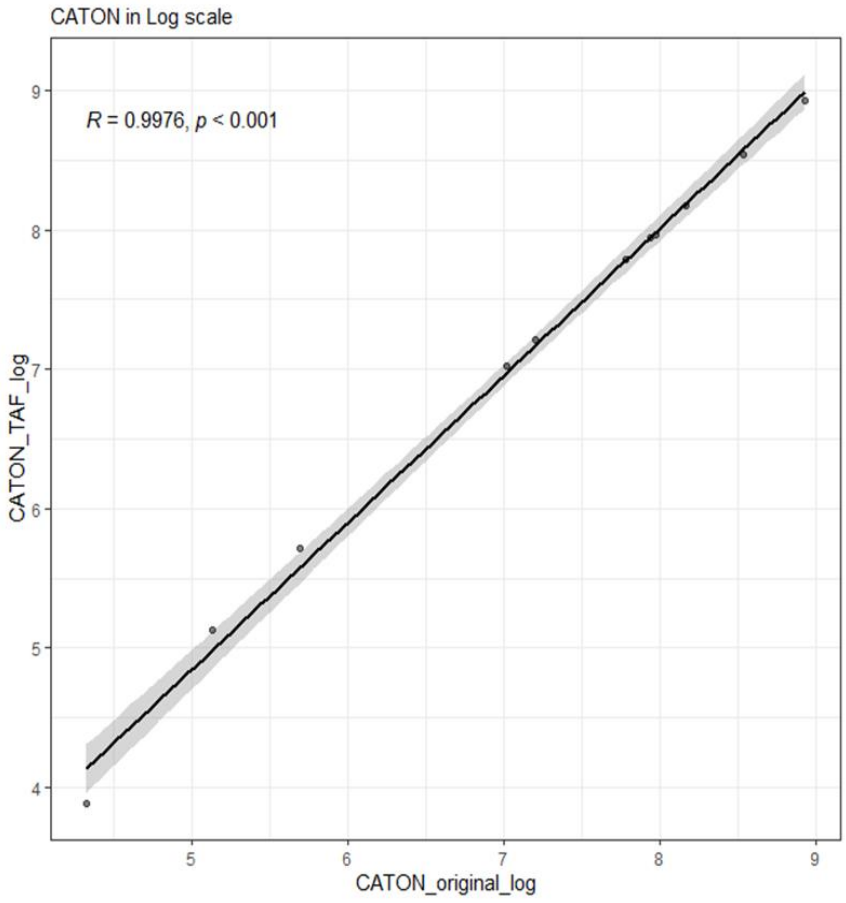
Overall we think that TAF would work ok with our present routines and support the collation of the relevant script for a stock quite fast, but we have some an issue with icesTAF::sourceAll(), since it doesn't support the way we work, see below

Issue

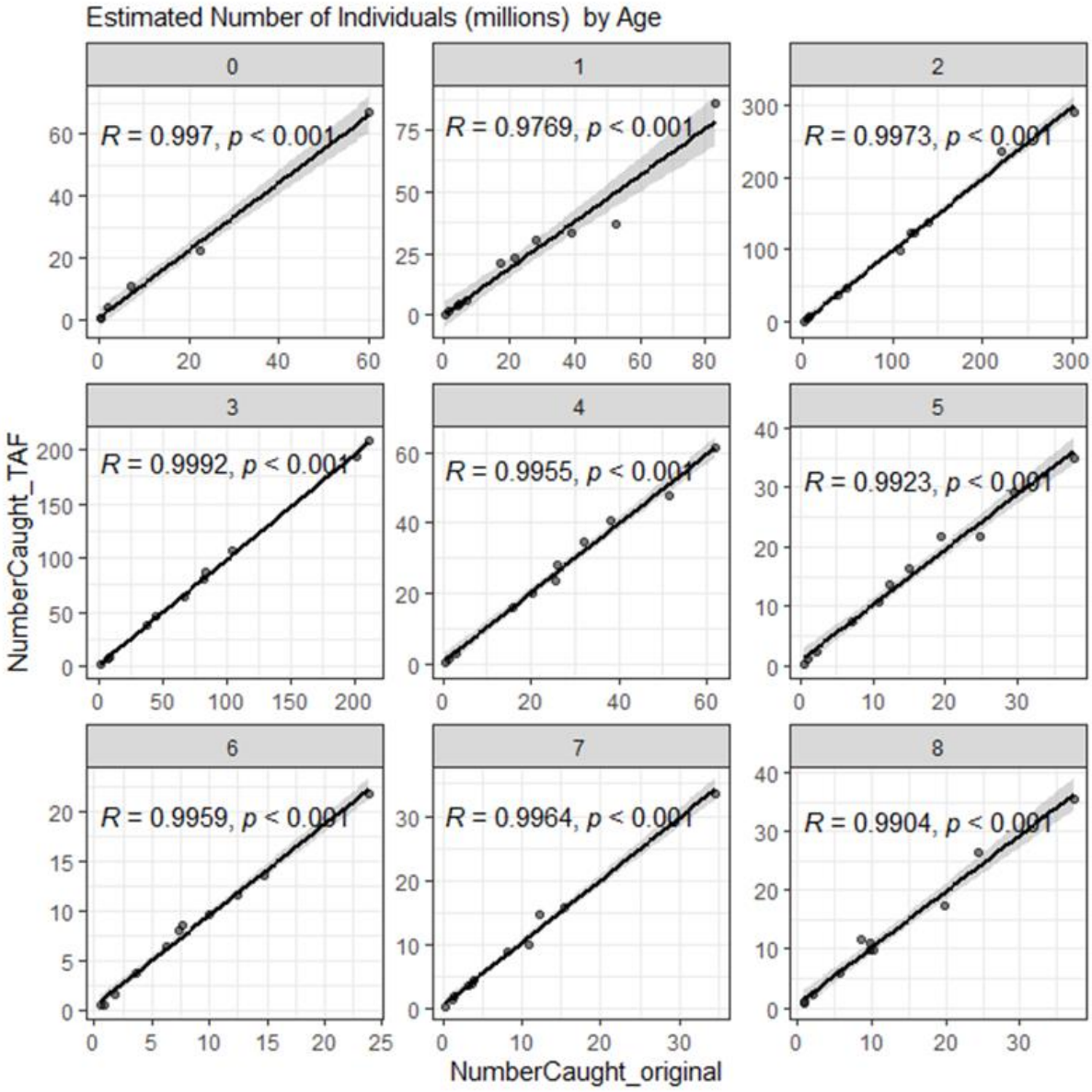
In general, we make the estimation per AWG and all data inputs (CL, CE and CS) are updated before we start estimating the relevant figures for an AWG deadline. Since all of the databases / data sets we use to create the RDBES data are open for updates all the time the likelihood of changes to data between AWGs are high, which means that rerunning all the estimation script with icesTAF::sourceAll() on our national repo would probably lead to minor differences in the outputs in our national repo and the ones in the SC repo. We assume that the intermediate database will be populated with an output script, if so then the data sets in the intermediate database would also be updated quite often during the year.

Estonia

We developed a TAF flow from RDBES national data downloads to Intercatch exchange files and compared the calculated national estimates with previously submitted InterCatch data. There is some variation between the obtained estimates. However, in general the correlation between the methods is fairly strong (R > 0.97 in all comparisons, see figures below). There seems to be a difference in assigning sampling trips and also landings into quarters though and some further processing of data that has not yet been incorporated into the TAF scripts. The TAF workflow is accessible at https://github.com/ices-taf/2023_all_RDBES_EE



Total catch in tonnes in log scale. The TAF calculated values are on the Y-axis. The original intercatch uploaded values are on the X-axis. Data is sprat (SPR) in 2022. Pearson correlation R is depicted on the graph.



Total estimated numbers caught by age class. The TAF calculated values are on the Y-axis. The original InterCatch uploaded values are on the X-axis. Data is sprat (SPR) in 2022. Pearson correlation R is depicted on the graph.

Finland

The goal was to reproduce the national LAN and LLN estimates for Baltic sprat, central Baltic herring and Bothnian herring stocks. The reproduction was started but is paused at this stage and is expected to continue and be completed in the following months. The main issue encountered is that the current data submitter - through InterCatch - for Finland was not available by the time of the workshop and following weeks, whereas the participant to the workshop is not a usual InterCatch user and the information he had at the time of the workshop was incomplete, which was noticed during the workshop. This issue will be fixed when both can meet again, and we expect to get the estimates reproduced by the end of the year.

France, Ifremer / MNHN (Museum of natural history)

We perform these raising exercises knowing that the RDBES submission was partial: CL discards, biological variables were not submitted. Moreover, metier6 of onshore sampling (H5) were not correctly computed. This restricts our analysis possibilities, so we decide to focus on LLN estimates. Two stocks were considered, syc.27.3a47d and sol.27.8ab, the first one during the week of the working group, the second one during the following week. One major issue in the RDBES data (catch registration) was fixed between these two studies and may contribute to explain some of the problems encountered in the first one.

syc.27.3a47d

Firstly, 2022 total landings for the stock syc.27.3a47d transmitted on InterCatch (IC) were compared at a metier6 level to the landing table CL transmitted via RDBES upload.

Secondly, reconstitution of the size structure sent on InterCatch for the same year and stocks were also compared at area level and metier5 level as the metier6 were not available on RDBES upload. For this comparison data needs to be raised the same way than it has been done for the InterCatch. Only landings data are used as discards data are not available on RDBES upload.

a) landings

For landings, total landings are similar for the stock with a difference lower than 1 kg (0.203 kg) between total landings from InterCatch and RDBEs upload, respectively 1,321,543 kg and 1,321,747 kg.

Results at metier6 level (Table 1) vary from 0.05 kg to 2,400.06 kg for similar metiers. Those slight differences must be explained by the new metier list developed for the RDBES.

IC metier6	IC landings (kg)	RDBES landings (kg)	Differences (IC-RDBES)
DRB_MOL_0_0_0_all	5468.52	5468.3	0.22
GNS_DEF_all_0_0_all	8161.08	8076.7	84.38
GTR_DEF_100-119_0_0_all	92196.93	92152.8	44.13
GTR_DEF_120-219_0_0_all	19237.21	19392.4	-155.19
GTR_DEF_90-99_0_0_all	160240.9	160240.5	0.4
GTR_DEF_all_0_0_all	4780.36	4624.3	156.06
LLS_DEF	8879.06	8960.2	-81.14
MIS_MIS_0_0_0	42261.34	40249.6	2011.74
OTB_DEF_<16_0_0_all	429.65	385.7	43.95
OTB_DEF_100-119_0_0	35469.9	36190.5	-720.6
OTB_DEF_32-69_0_0	7000.96	7632.3	-631.34
OTB_DEF_70-99_0_0	751496.35	751184	312.35
OTB_SPF_70-99_0_0_all	35483.71	35141.2	342.51
OTM_DEF	5176.73	5353.8	-177.07
OTM_DEF_70-99_0_0_all	3041.55	3041.6	-0.05
OTM_SPF_70-99_0_0_all	21130.31	20984.8	145.51
SSC_DEF_70-99_0_0_all	82422.94	84823	-2400.06
SSC_DEF_All_0_0_All	19536.88	19536.8	0.08
TBB_DEF_70-99_0_0_all	19129.11	18308.9	820.21

a) Length structure (LLN)

Results for the size structure did not work in terms of number of individuals raised. For several reasons, the estimation shows huge differences in the number of raised individuals between IC and RDBES upload respectively 1,671,083 and 745,279. Size structure patterns do not differ much per area and metier5 and could be explain by the metier5 aggregation but the scale on number of individuals raised are different (Figure 5).

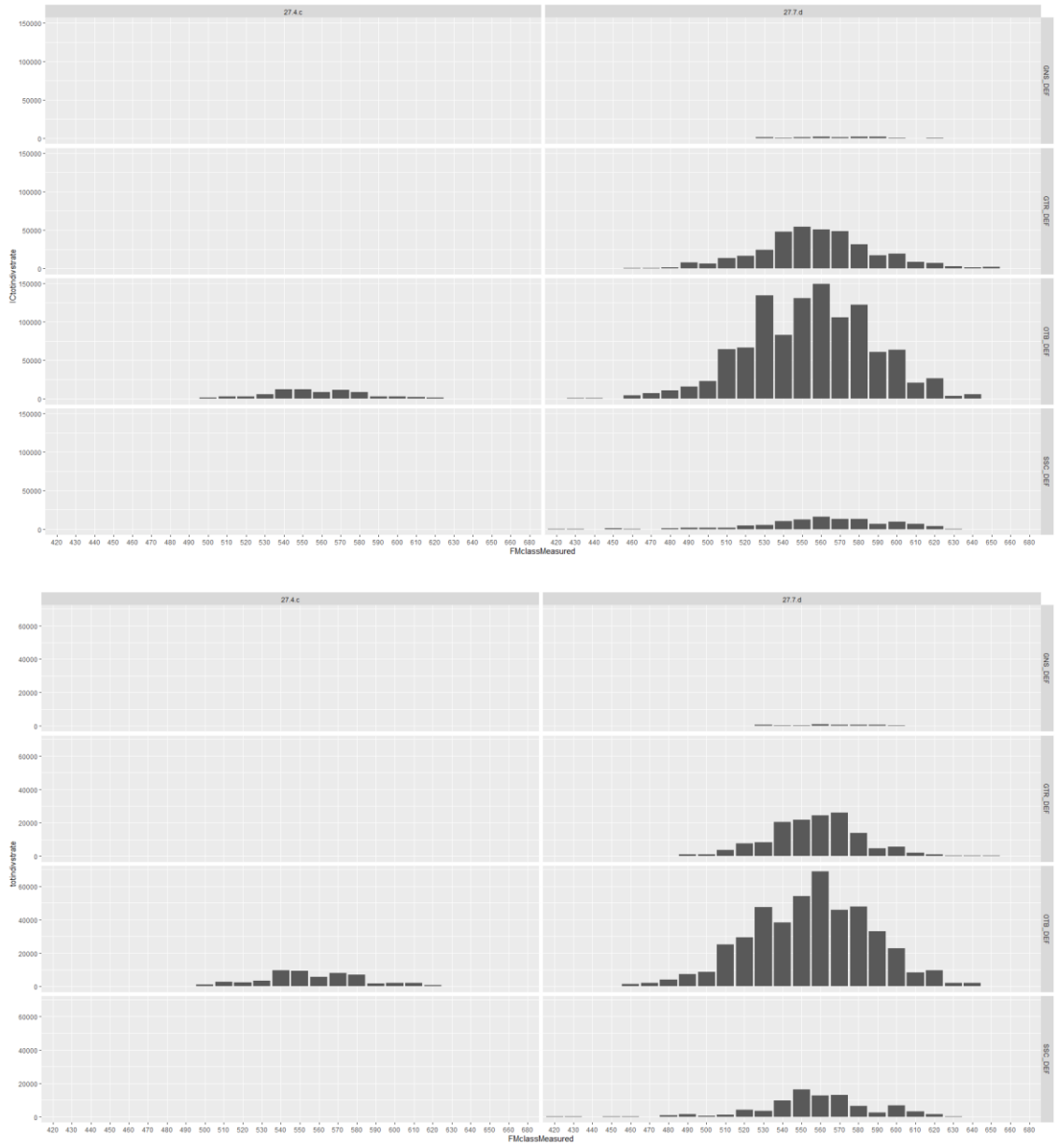


Figure 5 Number of individuals raised per size class by metiers and areas for IC data (above) and RDBES (below).

Potential failure reasons

1. Metier level

Obviously as the metier 6 are not available on the RDBES results could not be similar but we expect minor differences between IC and RDBES such as observed for landings.

2. Data cleaning and thresholds

Total individuals measured on IC and RDBES upload were respectively 2,787 and 3,363. This can be easily explained by thresholds applied to the data before to send the estimation to IC. Those thresholds are based on metier6 and area taking into account the number of samples as well as the number of length measured but also on historical landings (over 15 years) per metier. We were not able to reproduce the same method on metier5 as the number of samples per metier5 were not available on IC and landings historical series not available on RDBES upload. To get

around this issue, a list of metier5 per area selected on IC was created to filter the data at this level on RDBES upload. Obviously, this method brings some bias as the aggregation per metier5 could not be similar. Nevertheless, as the main metier were selected on IC the bias should not have induced such differences as the one found on raised numbers of individuals.

3. Estimation procedure

The main reason which could explain the differences lay in the raising process. Indeed, the person in charge of the raising for France was not available for the WK and could not explain the code used for the raising. Data were collected within two programs onshore at auction and the other one at sea on fishing vessels. This implies two different raising methods to access an estimation at a trip level. Consequently, reproducing the procedures turn out to be a challenge and differences in the number of individuals raised are highly probable linked to those steps on estimation work based on RDBES upload.

4. Data reported in RDBES column

Errors have been detected on information filled in some columns. This is the case in the column CatchRegistration which could also explain the differences in estimations.

sol.27.8ab

As a second example, we consider the ICES stock sol.27.8ab. For this particular stock, the mapping of IC fleets does not rely on metier6, but on gear and vessel length. We were so able to perform this mapping using RDBES data, despite the known problem of metier6 in the RDBES hierarchy V. We also solve the issue regarding the catch registration field (table FO, hierarchy I) before doing this exercise.

Since the French national raising procedure and sample selection include many validations steps, and given the absence of the person in charge of producing the IC data, we decide to reproduce raised length frequencies as follows:

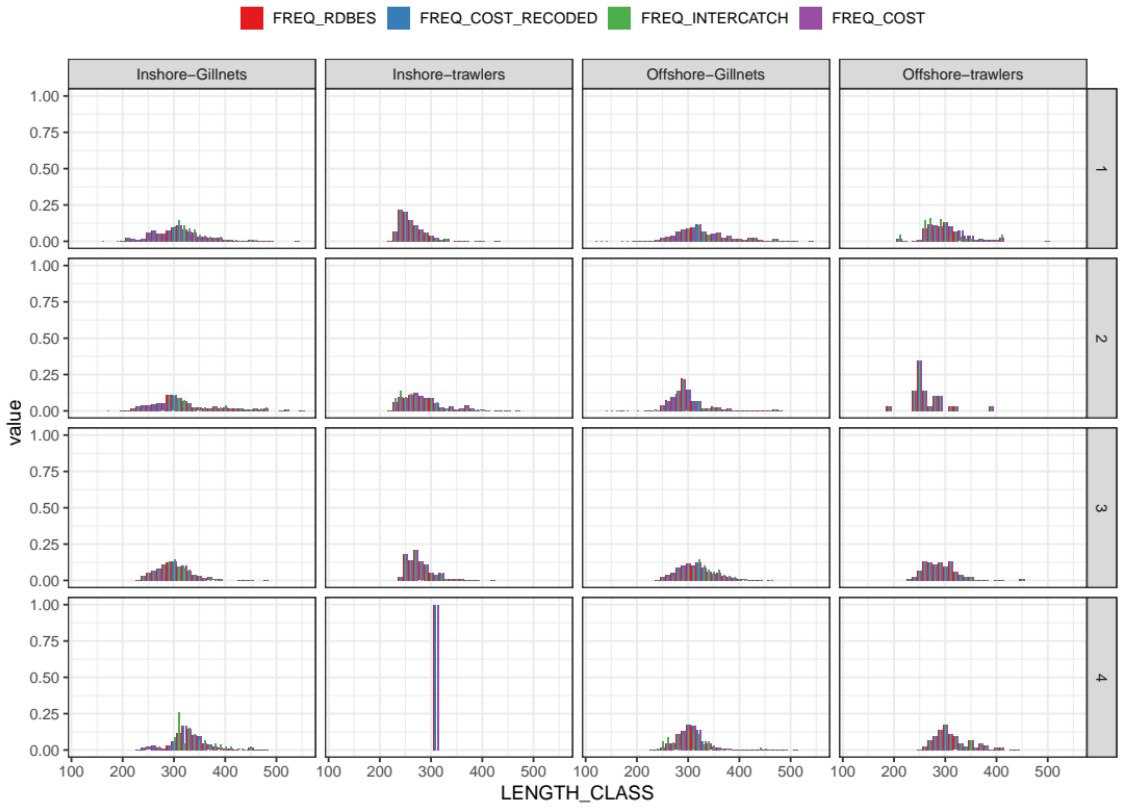
a) we first use the cost package and the cost dataset used for producing IC data to reproduce our own InterCatch estimates.
b) we recode the cost raising function to ensure a correct understanding of its raising algorithm.
c) we code the same raising function taking this time RDBES data as input.

Results for the test area 27.8b are shown on the following graph, while sampling information for each strata are summarised in the above table. We regularly, but not systematically, observe differences between IC length frequencies and our own raising. This is in line with the differences in the sampling data used for both estimates: selection of observation programs as well as sample validation performed for IC estimates are probably the main reasons explaining these differences.

The re-coded cost function always gives exactly the same result as our estimates using the cost function. Finally, the raising using RDBES data gives the same results, except for Offshore-Gillnets in season 2 where we can observe slight differences. An in-depth analysis has to be conducted to understand this difference.

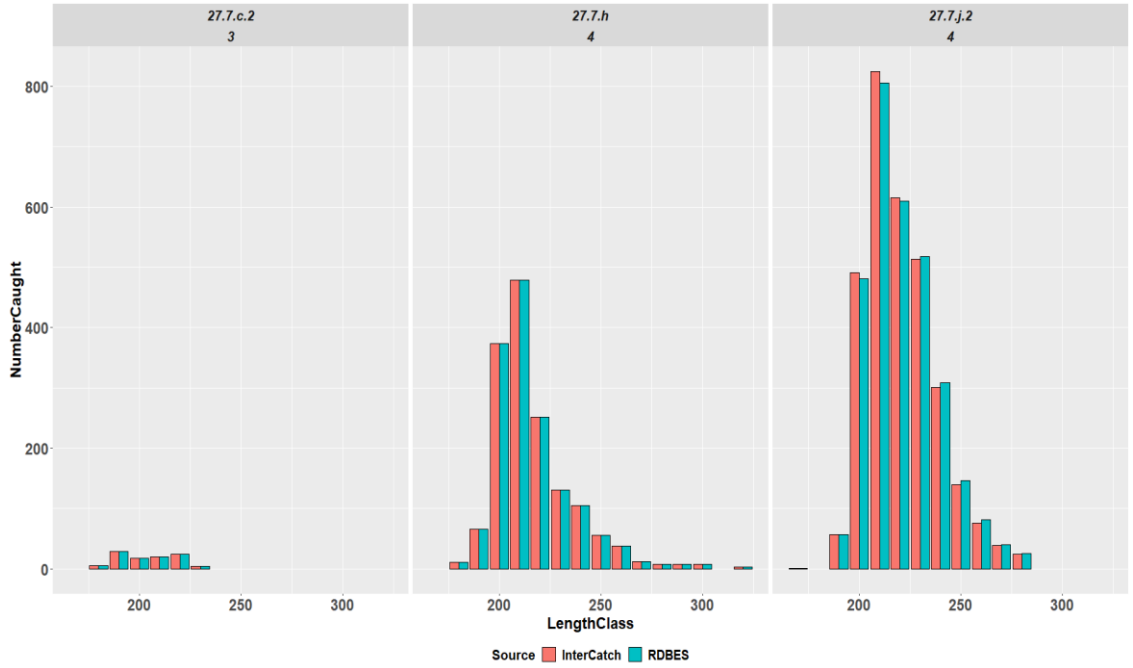
To conclude, this means that we should be able to reproduce IC LLN as soon as we know the exact sub-setting rules of sampling data. Additional studies with other stocks are required to test raising on sexed populations, and on discards.

Season	Fleet	FishingArea	IC SAMPLES	IC MEASURES	RDBES SAMPLES	RDBES MEASURES
1	Inshore-Gillnets	27.8.b	30	609	32	851
1	Inshore-trawlers	27.8.b	2	396	2	396
1	Offshore-Gillnets	27.8.b	20	824	20	824
1	Offshore-trawlers	27.8.b	4	201	4	201
2	Inshore-Gillnets	27.8.b	40	717	42	790
2	Inshore-trawlers	27.8.b	9	315	14	486
2	Offshore-Gillnets	27.8.b	16	800	16	800
2	Offshore-trawlers	27.8.b	3	29	3	29
3	Inshore-Gillnets	27.8.b	17	355	18	479
3	Inshore-trawlers	27.8.b	15	282	15	282
3	Offshore-Gillnets	27.8.b	6	668	7	1051
3	Offshore-trawlers	27.8.b	5	272	5	272
4	Inshore-Gillnets	27.8.b	3	46	4	283
4	Offshore-Gillnets	27.8.b	7	379	10	649
4	Offshore-trawlers	27.8.b	2	152	2	152
4	Inshore-trawlers	27.8.b			1	1

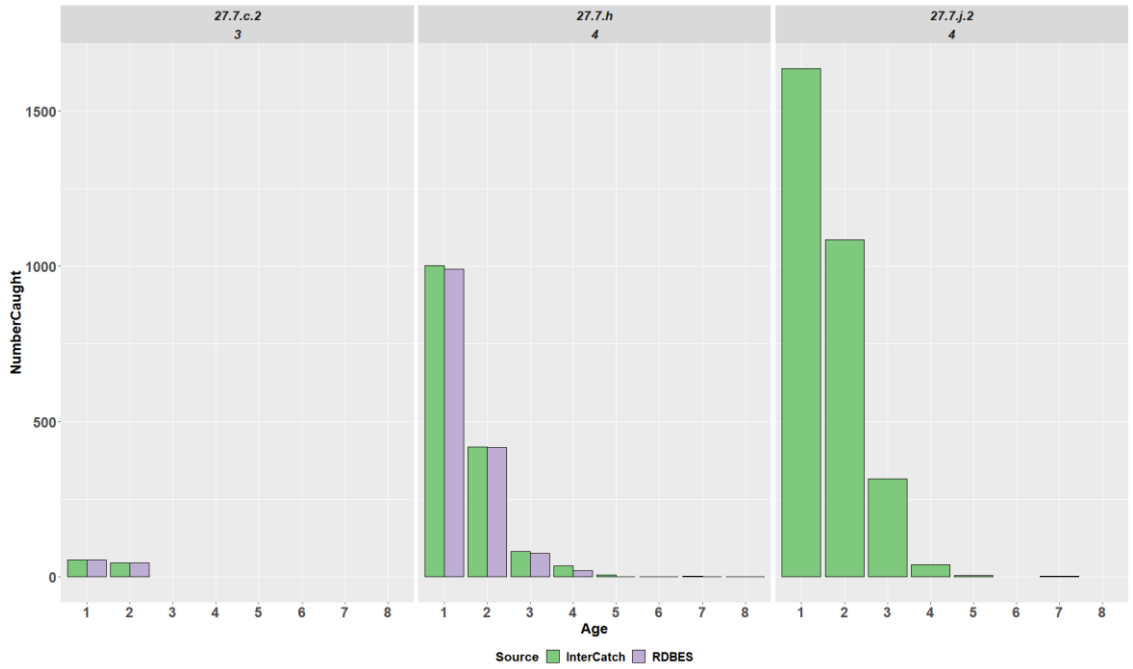


Germany

Our task during this week was to reproduce LLN and LAN estimates for selected stock (whb.27.1-91214), to compare them to those submitted to InterCatch (datacall WGWIDE). The corresponding length and age frequency histograms per area/quarter are presented below.



The length histogram reveals only minor differences in numbers at length for Intercatch and RDBES.



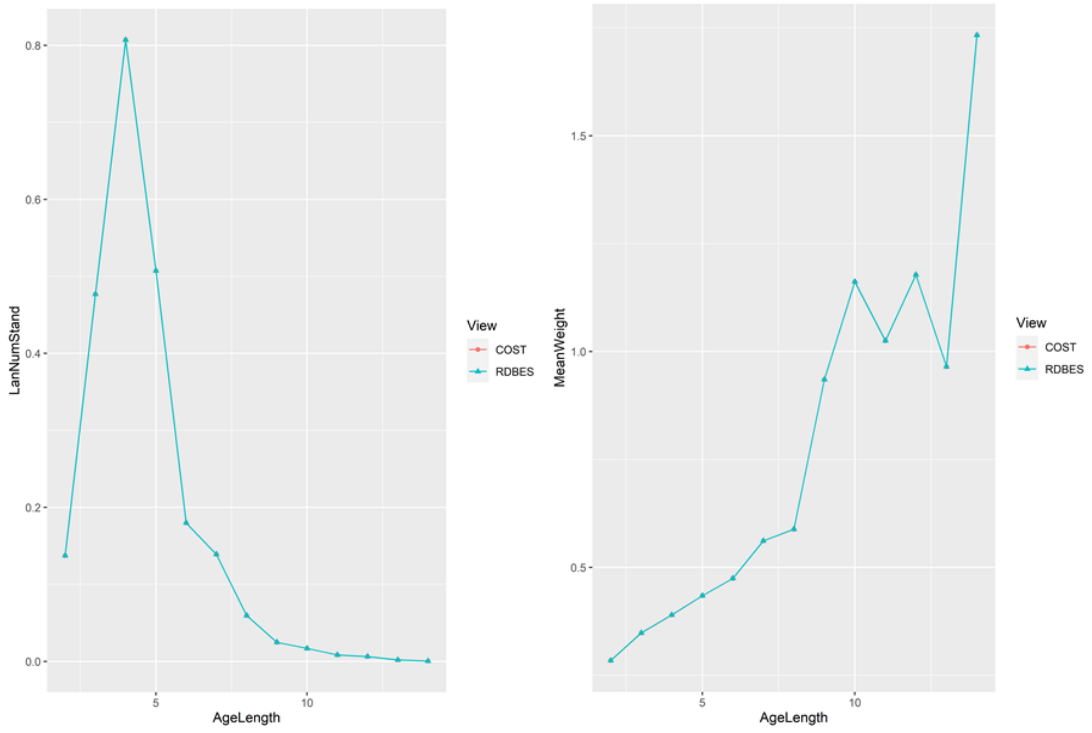
The age histogram demonstrates that RDBES age data are missing completely for area 27.7.j.2 in 4th quarter. But this can be easily explained: R estimation routine for InterCatch implements a procedure filling the gaps in sampling data, so missing ALKs for area 27.7.j.2 in 4th quarter were borrowed. That is not the case for RDBES, because only samples available in reality were used.

Ireland

ple.27.7h-k

The national estimation process for the length composition of the landings uses an adaptation of the code developed in the COST project (https://wwz.ifremer.fr/cost/). The approach for the workshop was to reproduce the COST input from data that was downloaded from the RDBEST and run the estimation procedures on this.

The figure below shows the outcome of the national estimation process using the current COST views and the views based on the RDBES data. For all length classes, the differences were less than 0.5%.



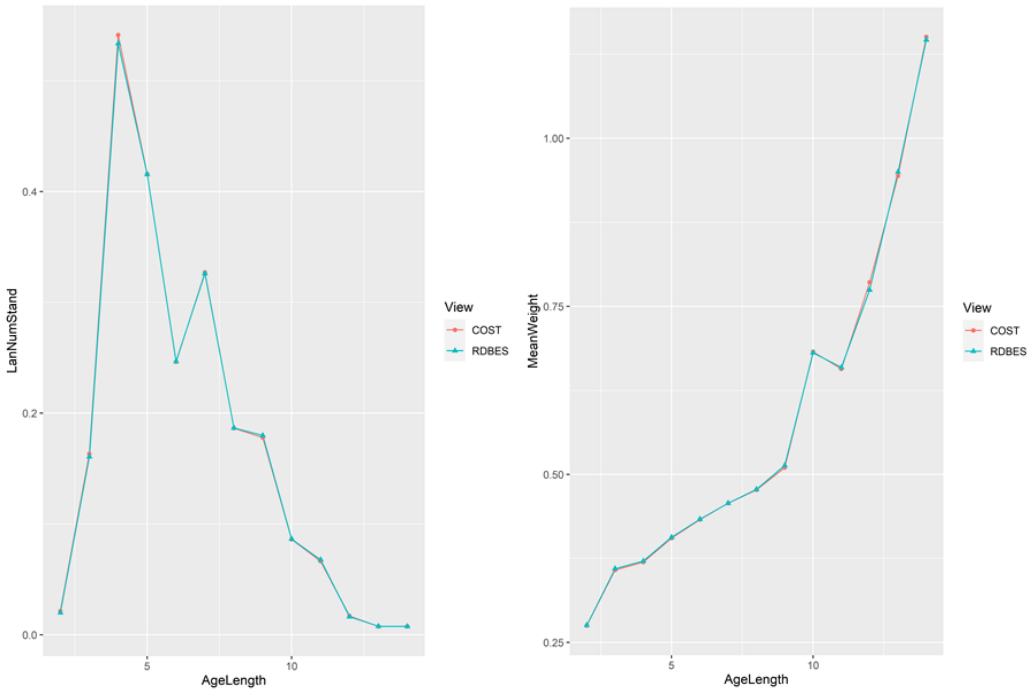
No gaps in the RDBES format were identified that prevented replicating the current estimation procedure.

The national estimation process for estimating discards volume and numbers-at-age/length is based on a set of scripts which could broadly be mirrored by scripts based on the RDBES format but more time needs to be spent on generalising these scripts and checking the outputs.

ple.27.7fg

The national estimation process for the length composition of the landings uses an adaptation of the code developed in the COST project (https://wwz.ifremer.fr/cost/). The approach for the workshop was to reproduce the COST input from data that was downloaded from the RDBEST and run the estimation procedures on this.

The figure below shows the outcome of the national estimation process using the current COST views and the views based on the RDBES data. For all length classes, the differences were less than 0.5%. These differences are explained by a difference in the underlying logbooks data.



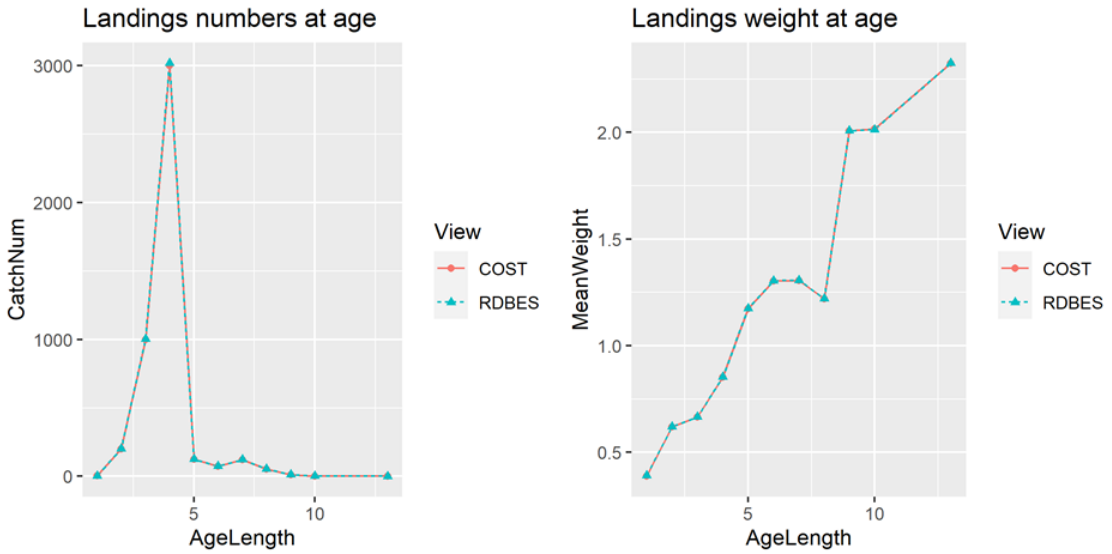
No gaps in the RDBES format were identified that prevented replicating the current estimation procedure.

had.27.7b-k

The national estimation process for the age composition of the landings uses an adaptation of the code developed in the COST project (https://wwz.ifremer.fr/cost/). The approach for the workshop was to reproduce the COST input from data that was downloaded from the RDBEST and run the estimation procedures on this. Minor differences were found and investigated, they included differences in the way metiers are assigned (or grouped); subsample raising factors; a small number of biological samples that were inadvertently removed.

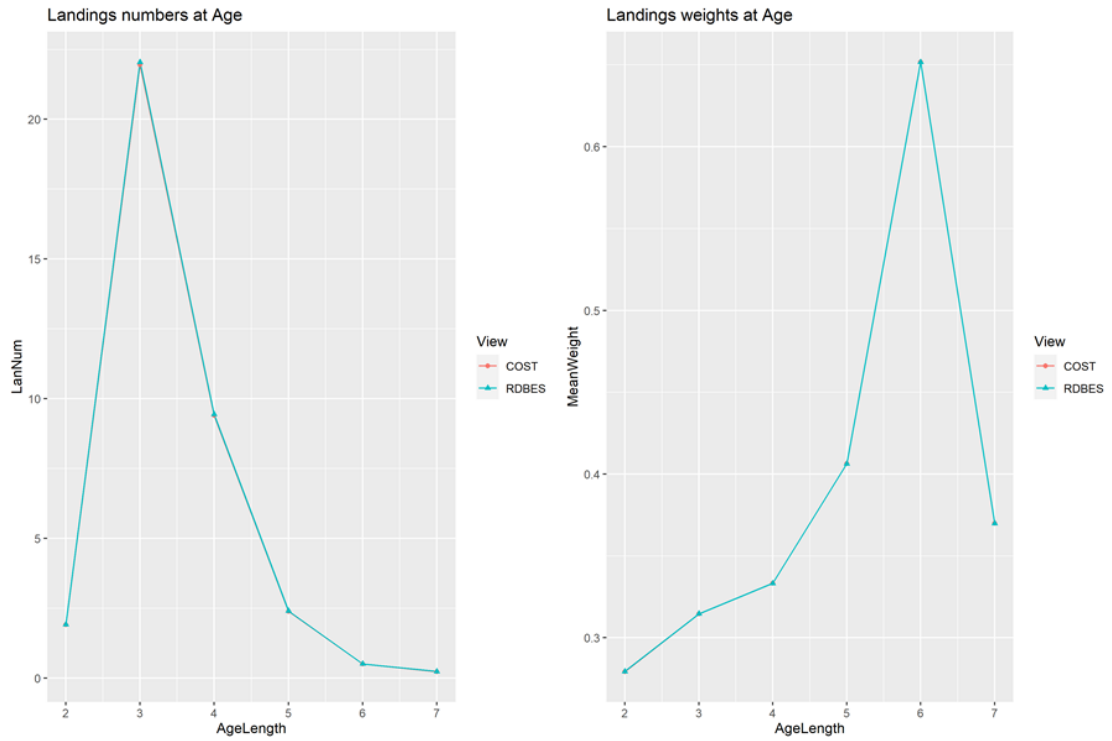
The figure below shows the outcome of the national estimation process using the current COST views and the views based on the RDBES data. For all age classes, the differences were less than 0.5%.

No gaps in the RDBES format were identified that prevented replicating the current estimation procedure.



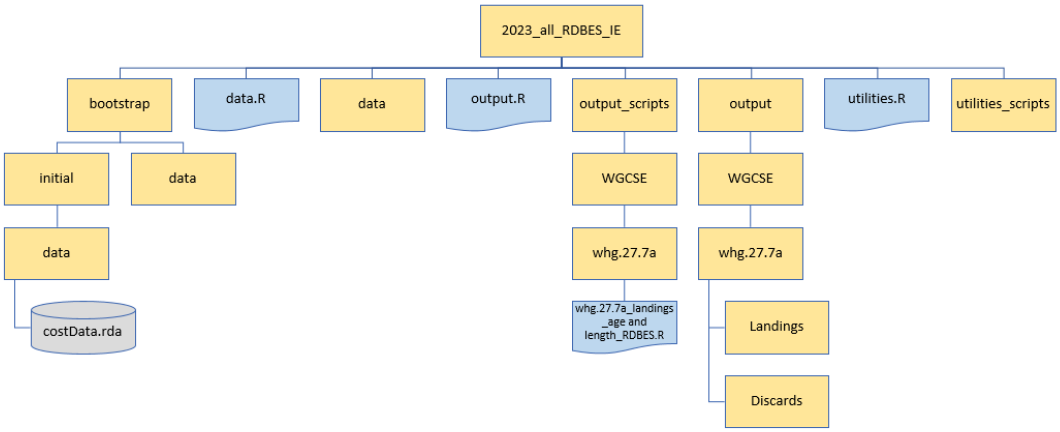
whg.27.7a

The national estimation process for the age composition of the landings uses an adaptation of the code developed in the COST project (https://wwz.ifremer.fr/cost/). The approach for the workshop was to reproduce the COST input from data that was downloaded from the RDBES and run the estimation procedures on this. Minor differences were found. The figure below shows the outcome of the national estimation process using the current COST views and the views based on the RDBES data. For all age classes, the differences were less than 0.5%.



TAF

The whg.27.7a landings scripts were migrated to a TAF repository and they were able to run successfully on a local PC. A system of sub-folders was created so that it would be easy to find the correct scripts and outputs for each assessment working group and stock in a national repository:



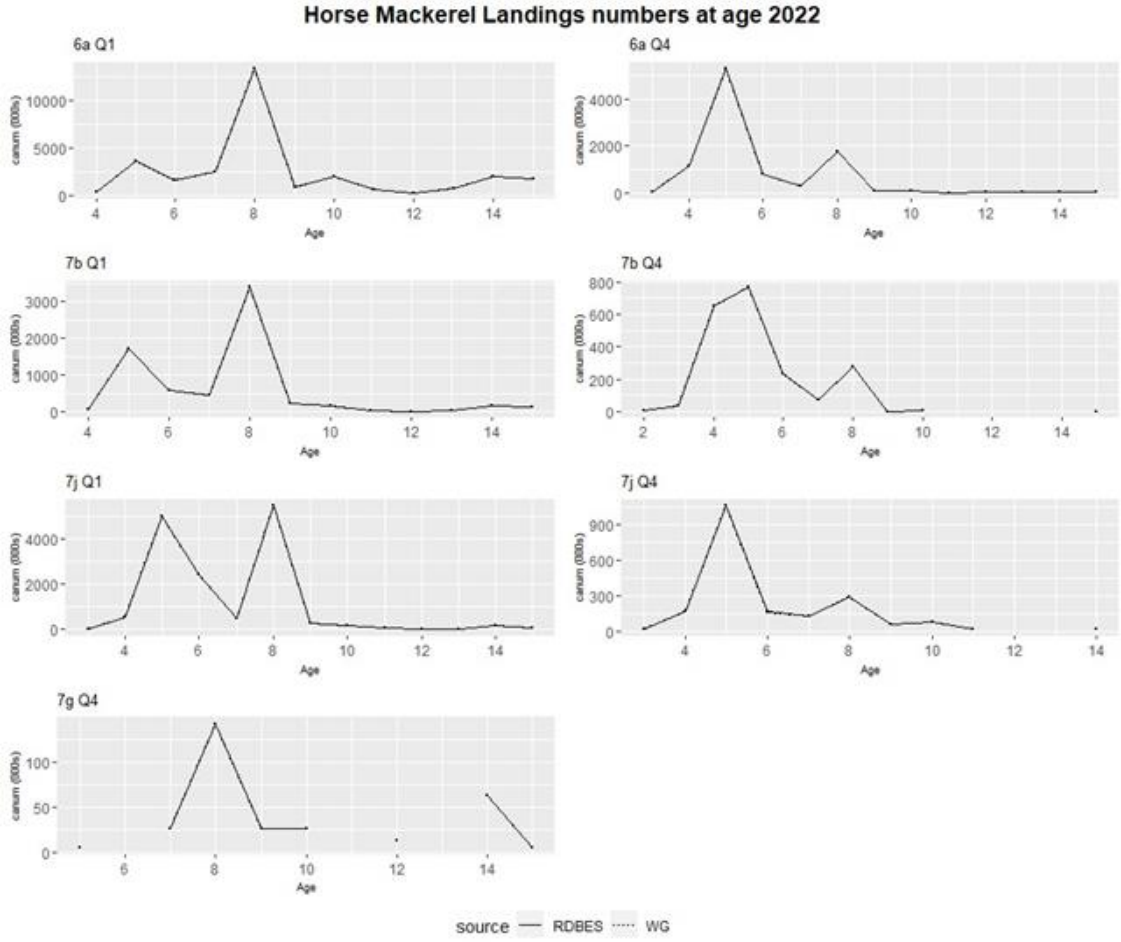
There are three important points to highlight:

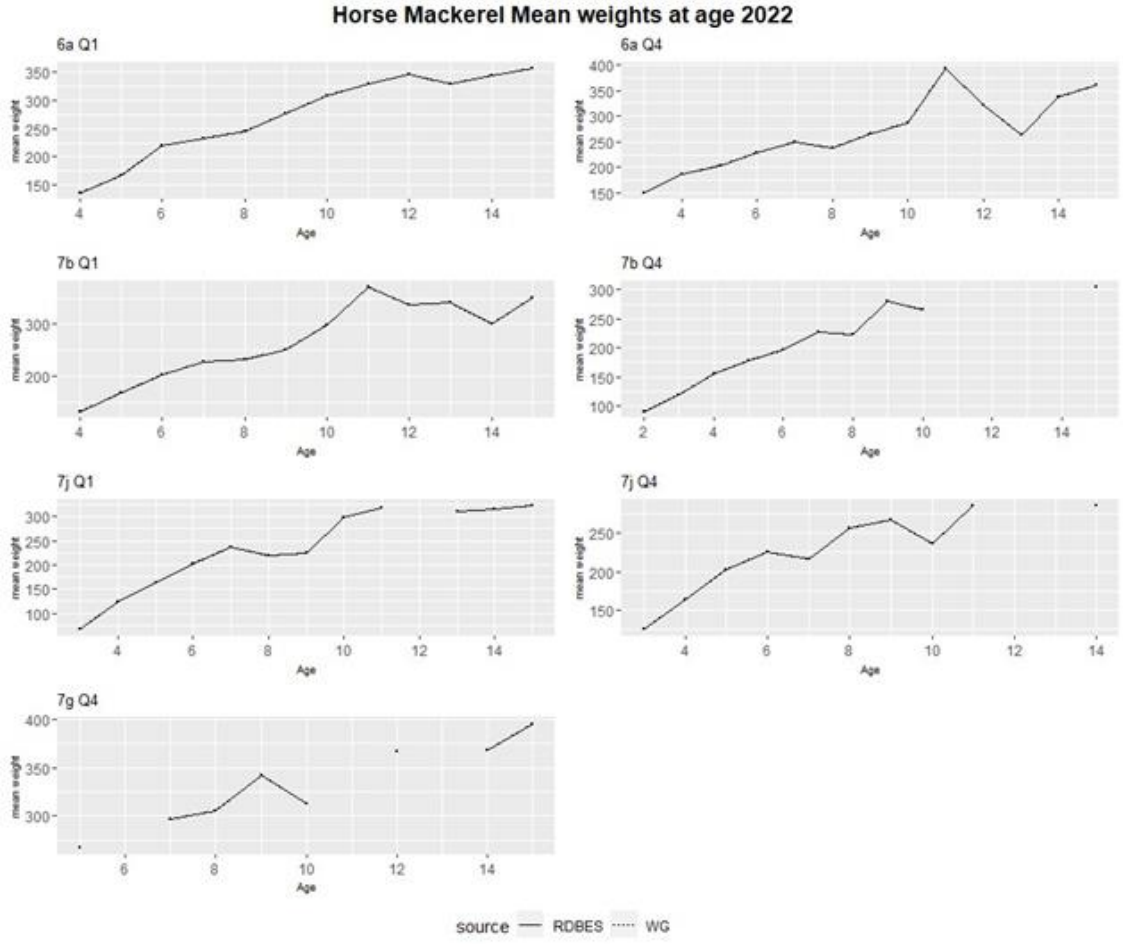
- The COST package does not work correctly in Rv4.x so we had to use an older version of R installed on the local PC - currently this is not possible on the TAF server.
- Our reproduction scripts rely on querying a national database to obtain RDBES data in the required COST format. These database views will need to be migrated to R script transformations.
- We will probably not be allowed to push RDBES data to a GitHub repository so currently we could not run the scripts on the actual TAF server.

hom.27.2a4a5b6a7a-ce-k8

The national estimation process for the age composition of pelagic landings for submission to the working groups uses a series of R scripts. The approach for this workshop was to reproduce estimates from data that was downloaded from the RDBES and run the estimation procedures using these R scripts. The landings numbers at age and mean weights at age comparing both methods are presented below. For horse mackerel in all ICES divisions, the differences were less than 0.5%.

No gaps in the RDBES format were identified that prevented replicating the current estimation procedure.

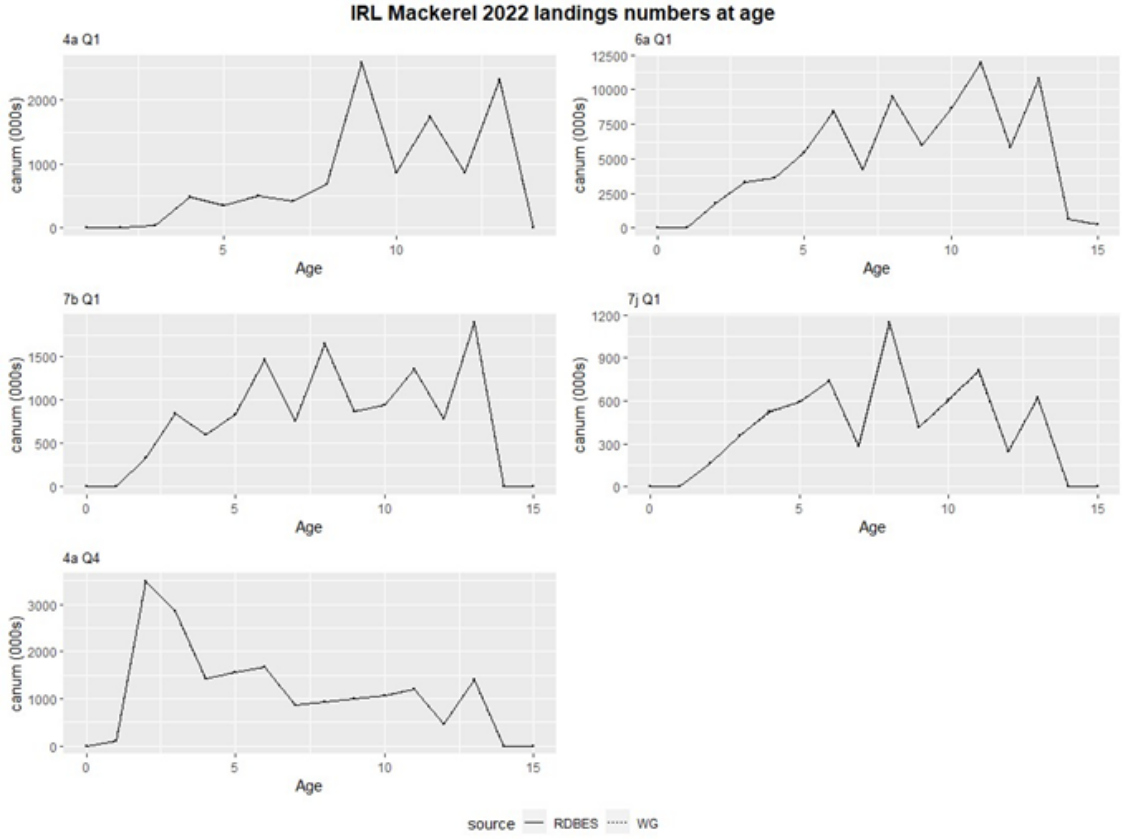

Horse Mackerel Landings numbers at age 2022
6a Q1
6a Q4
7b Q1
7b Q4
7j Q1
7j Q4
7g Q4
canum (000s)
Age
source RDBES WG

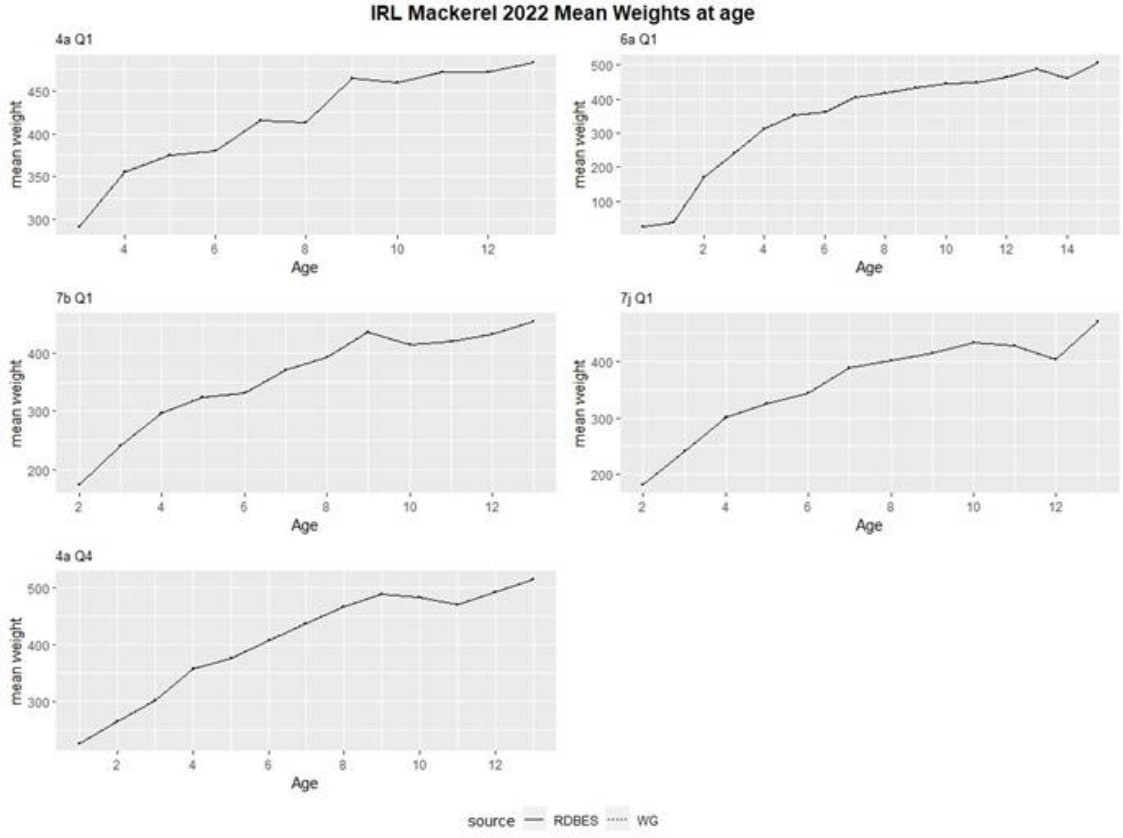


mac.27.nea

The national estimation process for the age composition of pelagic landings for submission to the working groups uses a series of R scripts. The approach for this workshop was to reproduce estimates from data that was downloaded from the RDBES and run the estimation procedures using these R scripts. The landings numbers at age and mean weights at age comparing both methods are presented below. For mackerel in all ICES divisions, the differences were less than 0.2%.

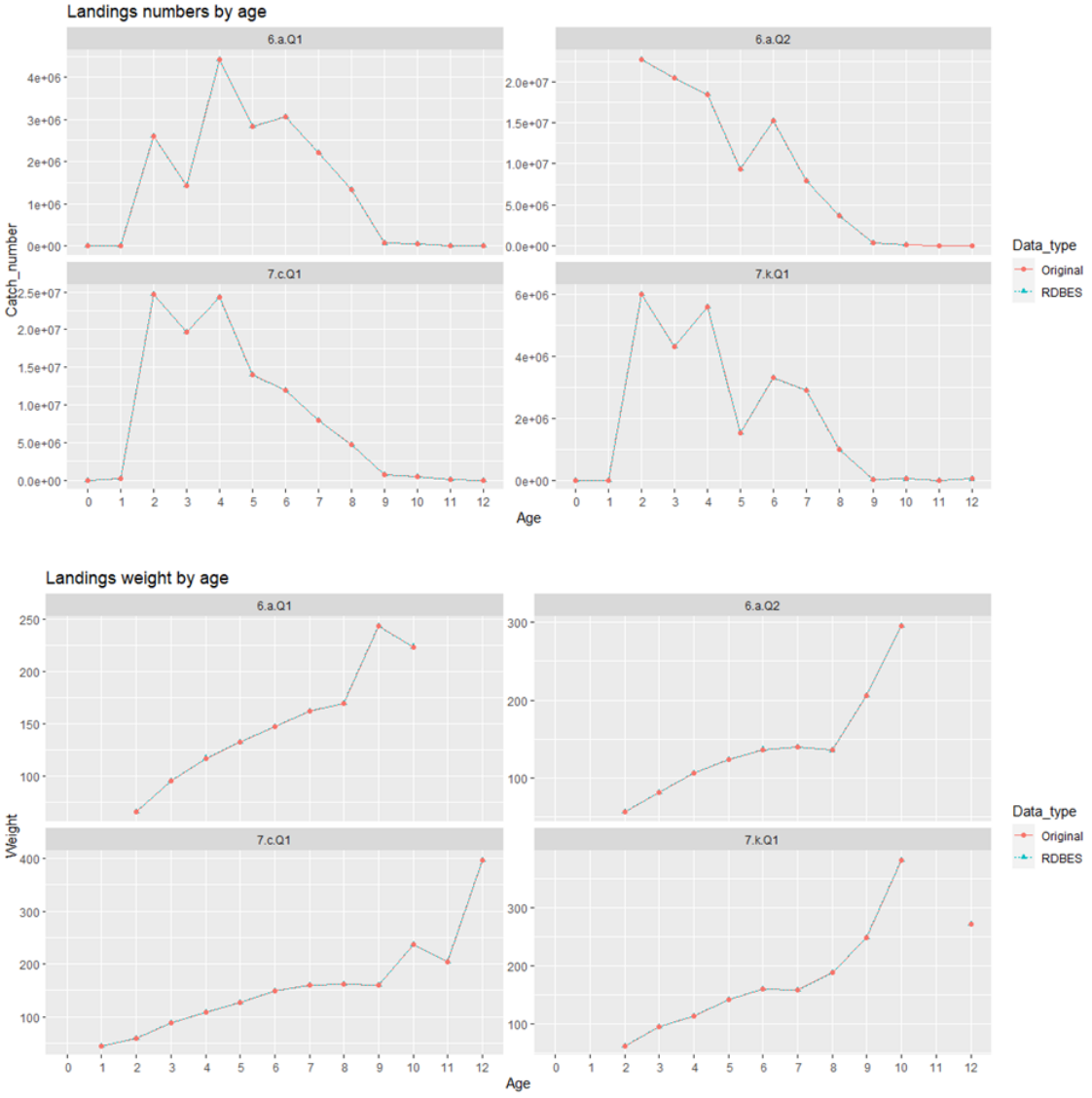
No gaps in the RDBES format were identified that prevented replicating the current estimation procedure.


IRL Mackerel 2022 landings numbers at age
4a Q1
6a Q1
7b Q1
7j Q1
4a Q4
canum (000s)
Age
source RDBES WG


IRL Mackerel 2022 Mean Weights at age
4a Q1
6a Q1
7b Q1
7j Q1
4a Q4
mean weight
Age
source RDBES WG

whb.27.1-91214

The national estimation files for blue whiting are produced through a series of R scripts which accessing the various databases and creates the InterCatch files for landings (by age and length) and discards. For this term of reference, data stored in the RDBES were used in the execution of the scripts and to reproduce the InterCatch files. The scripts ran successively and the files for landings by length and age were reproduced with very small differences observed (less than 0.001%).

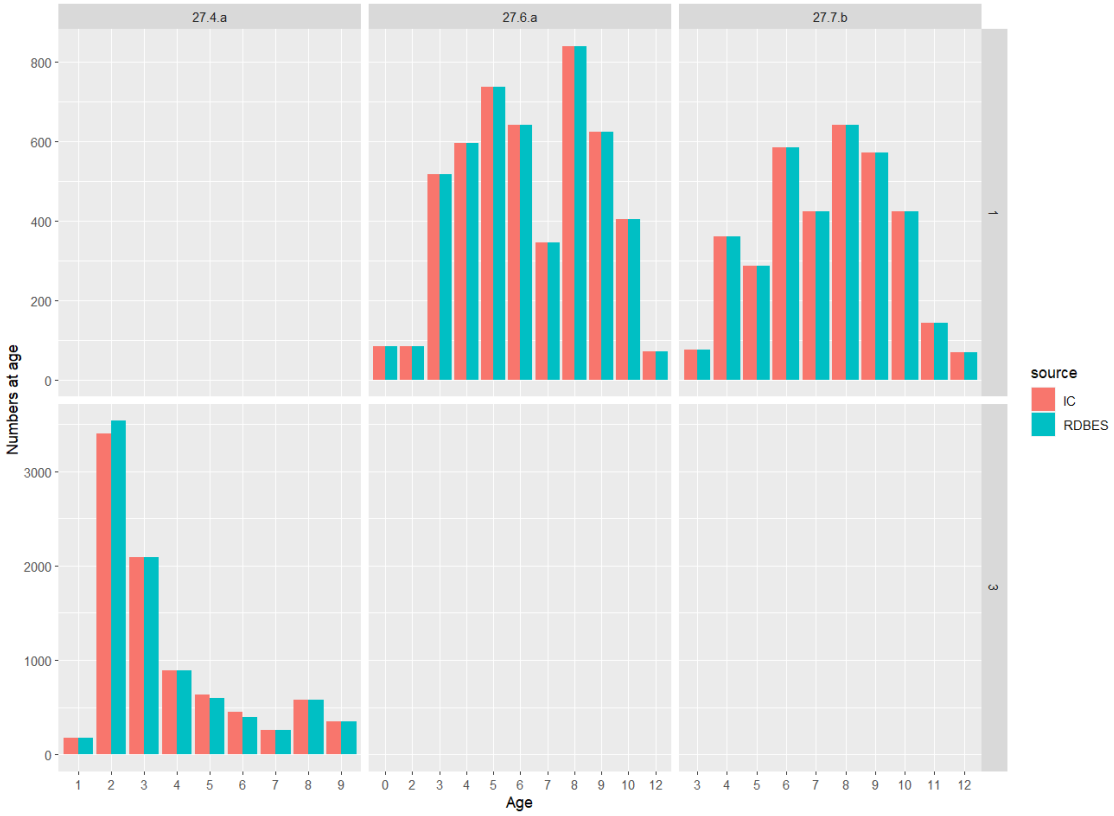


Netherlands

The goal was to reproduce the LAN and LLN estimates for the mac.27.nea stock. Successful reproduction ensures that it is possible to reproduce these estimates for all the pelagic stocks the Netherlands reports to ICES. LAN estimates were reproduced successfully and no issue was identified with the RDBES data model. Minor differences are attributed to the data upload. LLN could not be reproduced, however, this is related to data omitted during the upload. Furthermore, it was possible to reproduce the weight landed by ICES "domain" corresponding to the InterCatch SI records.

TAF

No substantial progress was made under TAF. However, it was identified that there is a need for further code development.



Norway

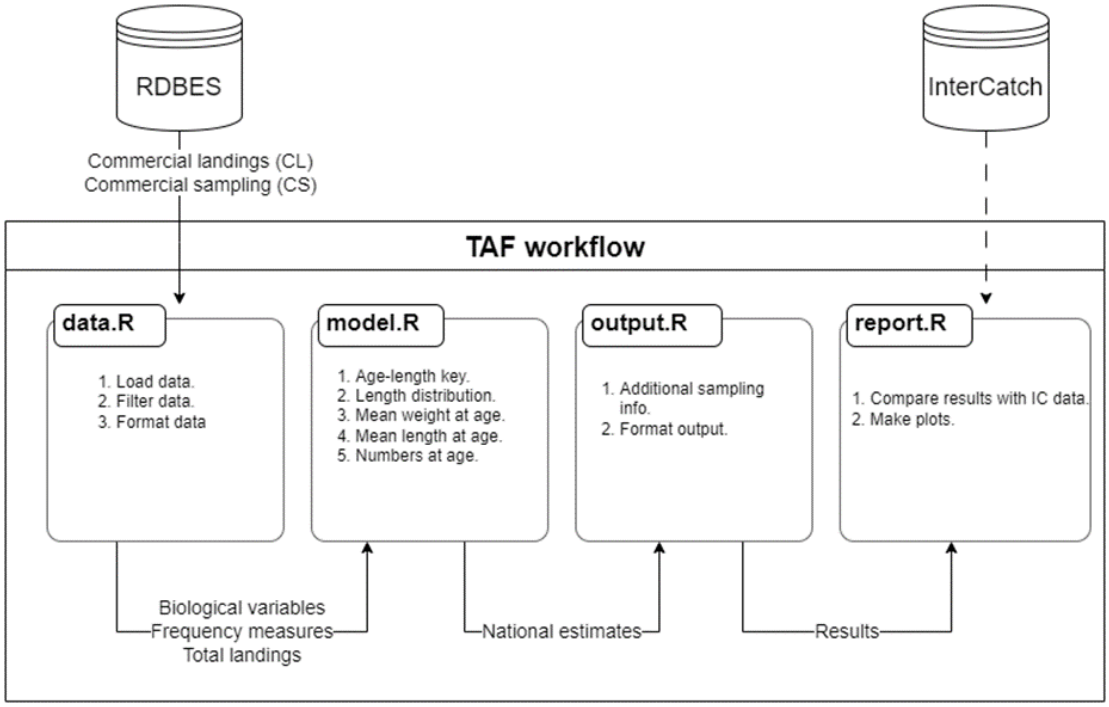
There was no additional progress in this work, as Norway did not submit data to RDBES in time for the workshop. From Raise&TAF 2022 workshop it was learned that a complete implementation and a partial reproduction could be made, and from the Raise&TAF_Flow 2022 workshop it was learned that a complete TAF-implementation could be made. We have reviewed if any changes to the Norwegian data flow or the RDBES or TAF would impede implementation and reproduction. None such issues have been identified.

Poland

During the WKRDBES-Raise&TAF2 workshop, NMFRI representatives focused on reproducing national estimates of the selected stock. The Baltic plaice stock (ple.27.24-32) was chosen for this task. At the previous edition of the workshop in 2022, the national estimation routines were successfully adapted to the RDBES data format. The main objective for this workshop was to convert the routines into TAF workflow.

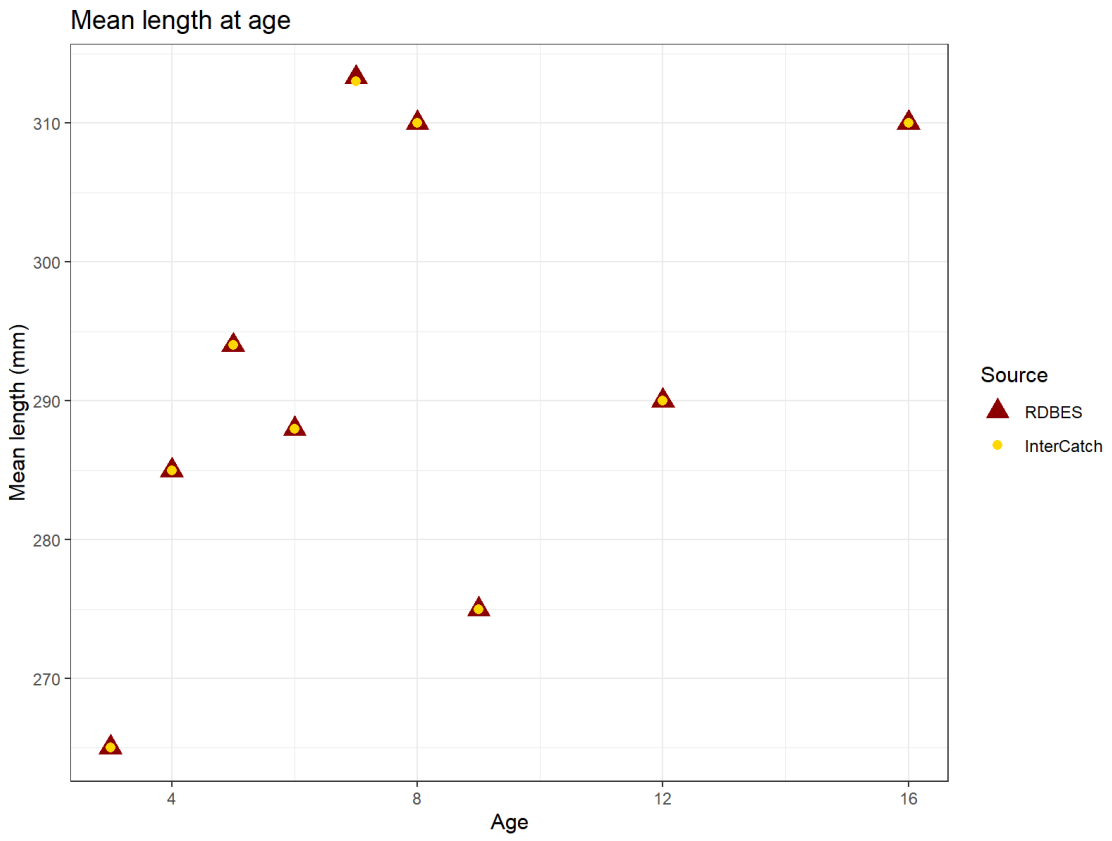
The latest version of commercial landings (CL) and commercial sampling (CS) data from 2022 were downloaded from RDBES. The data were drafted and referenced in the TAF environment. Additionally, the data downloaded from the InterCatch were included in the analysis, to check if the reproduction of estimates was successful.

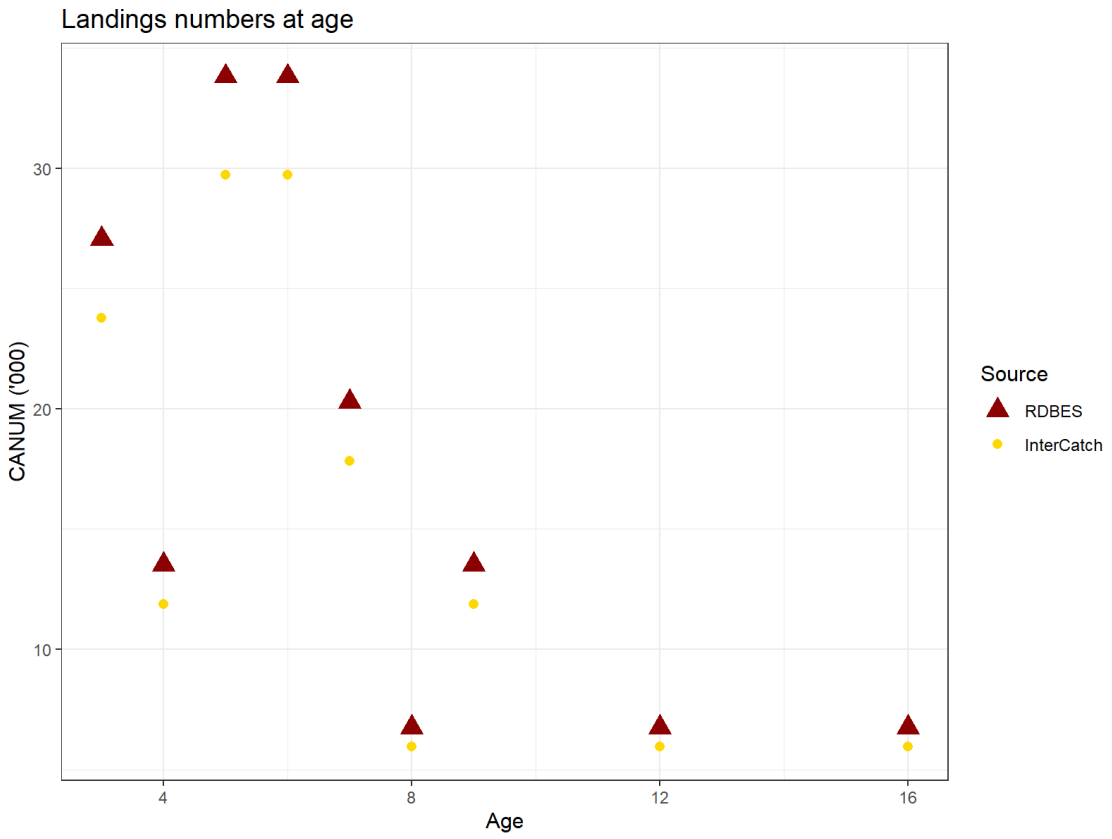
The repository for the Polish national estimation procedures was set up by ICES on GitHub. The scripts were developed following the TAF guidelines provided by ICES. The diagram below shows the flow of data between different modules of the repository.



The reproduction of national estimates in TAF was partially successful. The reproduced mean weight at age and mean length at age fully match with the values submitted to InterCatch. There is a difference in numbers at age. However, this issue is related to the updates of official landings statistics, so this is not a limitation of RDBES data model. The plots below show the comparison of the reproduced values with the ones in the InterCatch.


Mean weight at age
WECA (kg)
0.30
0.27
0.24
0.21
4
8
12
16
Age
Source
RDBES
InterCatch


Mean length at age
Mean length (mm)
310
300
290
280
270
4
8
12
16
Age
Source
RDBES
InterCatch



Spain, AZTI

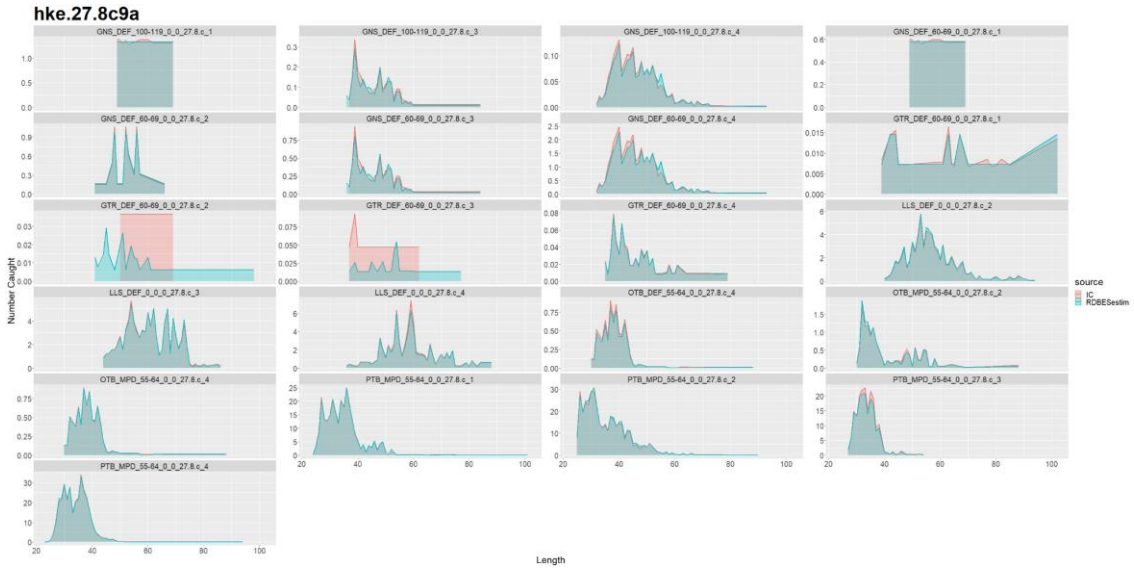
LLN estimates

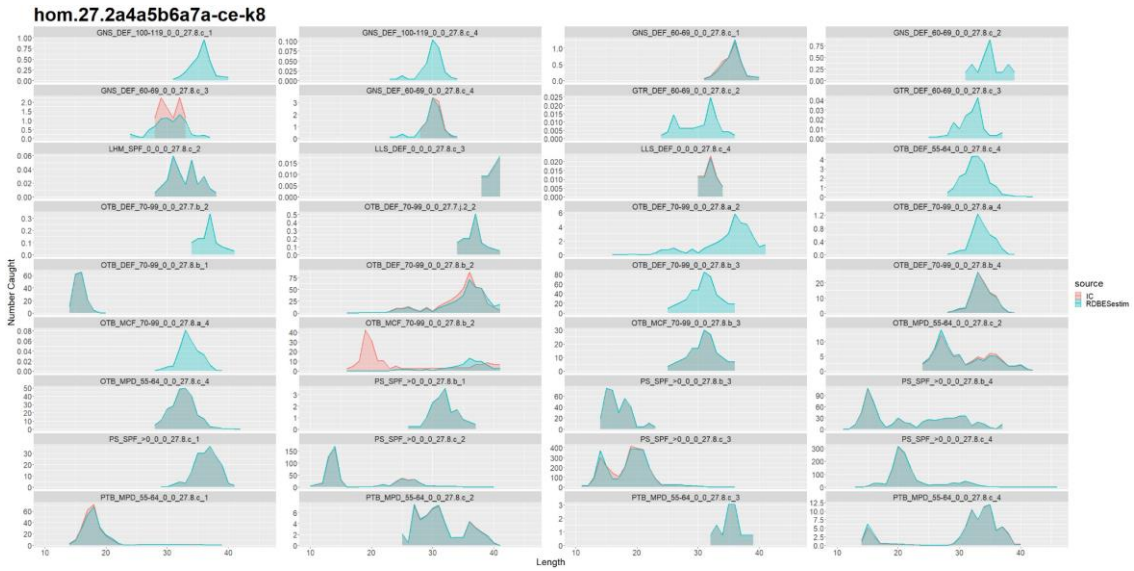
We made progress in our scripts to reproduce LLN estimates for several stocks (H5), but we could not complete the reproduction. The problems encountered were the following:

- CL table for year 2022 was not available.
- Using CL table from 2021, we realised that we cannot distinguish the landings sampled by IEO from those landings sampled by AZTI (each institute is using a different sampling scheme within the same H5). We need to talk with the Spanish Administration to find a solution. We don't need changes in the data model, but we need some of the optional fields to be completed. We see two possible solutions:
    - To fill in the variable CLsamplingScheme. In our national Databases we have this information trip by trip, indicating which Institute is in charge of the sampling. This option is our preferred solution and should give exactly the same results that we are providing to IC.
    - To fill in the variable CLlandingLocation, without using the dummy code "ES999". Missing values can be filed with the Arrival Location/ DepartureLocation/ RegistrationPort. This option is a proxy, although it will not give exactly the same results that we are providing to IC.
- We achieved quite good results in the reproduction of LLN of the stocks where we are not making allocations (i.e., hke.8c9a). "No allocations" mean for us that LLN estimates are not calculated in the strata without sampling. However, we still need more work to

reproduce those stocks where allocations are made (i.e., HOM). See figure below. These results apply only to the estimations made by AZTI.

- The exercise of the two institutes working together in the reproduction of the RDBES data based estimates has been useful, as some "hidden" differences in the process have popped out. We are trying to reproduce the way we are working now, including these differences.
- The Spanish estimates are made by summing up the estimations made by IEO and AZTI. We didn't have time to compare the Spanish estimations using RDBES data and the final IC file sent to ICES. This work is in progress.





DLN estimates

We started developing the scripts for DLN estimates (H3). We haven't been able to reproduce the estimates yet. So far, the main problems encountered are the following:

- In trawls trips sampled on board, we always have some hauls which could not be sampled (the observers need to rest). In our Database, we apply expert knowledge to non-sampled hauls, and assign them the species composition and length distribution of other sampled hauls in the same trawl trip. We need to find the best way to reproduce those imputations using RDBES data, but the results will not be exactly the same.
- In the discards estimates we are assigning all hauls of a trip to the dominant area of that trip. At the moment we cannot report the information of the dominant area in H3, but we have posted an issue in GitHub and already have some ideas to solve the problem.

TAF

We haven't started using TAF in the reproduction. We preferred to focus on the development of the scripts to calculate the estimates and also to coordinate it with the IEO. We plan to start using TAF standards when the calculation of estimates is in a more advanced state.

Spain, IEO

LLN

As data submitters we have been working on the reproduction of national estimates for H5.

- Since no data from CE and CL was available in the RDBES database for 2022, we employed the data from 2021.
- We have selected a simple case study selecting hke.27.8.c.9.a. This task was selected in collaboration with AZTI to aggregate later both results (IEO+AZTI) to compare them with the data submitted to the working group (WGBIE).
- No major problems with the estimates of this stock were encountered and the data from RDBES seems to be enough to complete the task. Nevertheless, this same task for other stocks could be more complicated in the case of some stocks where the onshore samplings are employed to distribute the official landed weight per species (e.g. HOM, MNZ, LDB, ANK, MON…).
- Since at laboratory IEO allocations are not performed to create IC files, our estimates can be easily replicated using RDBES data.

DLN

We have not started estimations for DLN (corresponding to H1 for laboratory IEO).

For 2022, data from H1 was not available at RDBES database, as was the case for other countries under H1.

TAF

None of the work performed has been done in TAF.

Sweden

Our aim for this workshop was to continue from where we left last year. Some participants focused on extracting commercial sampling data from RDBES and comparing it with the full 2022 dataset used for data calls in 2023. Other participants focussed on trying to replicate earlier estimates. The participants extracting data proceeded significantly in extracting data compared to work done in 2022 when we had problems in fully understanding the use of the functions. The updated functions in the new library RDBEScore worked smoother compared to the previous library (icesRDBES) used during Raise and Taf in 2022. A number of scripts were constructed

during and after WKRDBES-Raise&TAF in 2022 to combine the RDBES data format into a format resembling the RDB format used in many of the current scripts used for estimation. These scripts were crucial to start the comparison of data in the new format with data in the old format that was done during Raise&Taf2.

An attempt was made to reproduce earlier discard estimates (total discards in weight), using the RDBES format (not transformed into the old RDB format). This was successful for most attempted stocks and strata. However, for some stocks/fisheries (mainly one self-sampling programme) auxiliary information not present in RDBES have been used previously, and more work is needed to decide how to deal with those.

Issues

- As the RDBES data format differs in many ways compared to the previous RDB format it took a lot of time just to compare the different datasets - actually much more time than anticipated.
- As sampling frame and strata grouping in the new format differs significantly from how it was in the old format, sub-setting datasets for different types of estimations was not as straightforward as anticipated.
- For some of the participants it would be a huge task to rewrite current R scripts to be applicable directly to RDBES extracted data and not via the recombination mentioned in the first paragraph.

TAF

We followed the developments in plenary but have not yet started work on the TAF structure.

UK (England)

The aim for this week was to continue to develop R code to reproduce the discards weight estimates and discards length compositions. As per last year, we have developed and reproduced the method to provide discards estimates and discards length compositions from the offshore sampling programme (H1). This year we had all the necessary data and completed fields to attempt to estimate the discards for the Celtic sea stocks. However, the estimations were significantly different from the ones submitted to InterCatch. Data exploration was carried out to understand the source of the differences. The identified issues are listed below:

- Missing sampled trips submitted to RDBES;
- In some instances, the calculation of the sample ratio estimator (SAnumTotal and SAnumSamp), was considerably different. The differences were due to a rounding issue, as the RDBES only accepts numerical to one decimal place (e.g. 0.25 rounded to 0.3);
- Different/Incorrect metiers allocations;
- Differences on how the data from the Self-sampling programme is dealt with for the Intercatch submission and for RDBES submission.

A general issue with the UK relates how the UK reports to ICES. Currently each Devolved administration (DA - England, Scotland, Northern Ireland, Wales) submits the data individually following the rule - UK vessels landings into the DA and DA vessels landing abroad. However the CL and CE tables access is granted for individual DA rather than for the whole UK. Therefore, to reproduce the InterCatch submissions, we have used the landings extracted from the national database, rather than CL or CE tables.

No progress was done on trying to put the code in TAF.

UK (Scotland)

UK reports landings to ICES stock assessment working groups in an unusual way for historic reasons. Essentially, each devolved administration (DA - England, Scotland, Northern Ireland, Wales) reports UK landings into DA and DA landings abroad, rather than DA landings everywhere as other countries would do. However, CL & CE access is granted for individual DA rather than for the whole UK.

In this workshop we intended to use the new auxiliary variable field to hold the relevant landed weights, however the sampling strata (fishing season) do not completely align with the estimation domains (quarter, area).

To reproduce the Scottish estimates of n-at-length for the stock mac.27.nea using RDBES data, we successfully connected to the ices TAF repository by following the instructions on Getting set up · ices-taf/doc Wiki · GitHub, and connection of GitHub account with ICES username (taf.ices.dk/github).

The 2023_all_RDBES_GB-SCT TAF repo was cloned as a project in RStudio. Push changes were not possible as no write access is provided to this TAF repository.

The data.r script was edited to read a locally stored zipped download of all Scottish RDBES CS data, using the function RDBEScore::importRDBESDownloadData. The structure of the imported RDBES data is a list of the CS tables (SA, FM, SS etc.). The function TAF::write.taf can only export single tables. Therefore, the script separately exports, to the data folder, each lower hierarchy table that will be required for estimation. It would be useful if the function TAF::write.taf could export all lower hierarchy tables in one action, rather than requiring each table to be exported separately.

In the script output.r, each lower hierarchy table is read into the workspace with the function TAF::read.taf. The function RDBEScore::filterRDBESDataObject did not correctly filter the data, so the data is manually filtered to the stock at this point. Using ratio estimation scripts developed during WKRATIO, the n-at-length were estimated. There was a difference of approximately 0.2% between the InterCatch and RDBES-based estimates for 2022 MAC data (Figure 6). Rounding and the use of a separate landings file may have contributed to the differences.

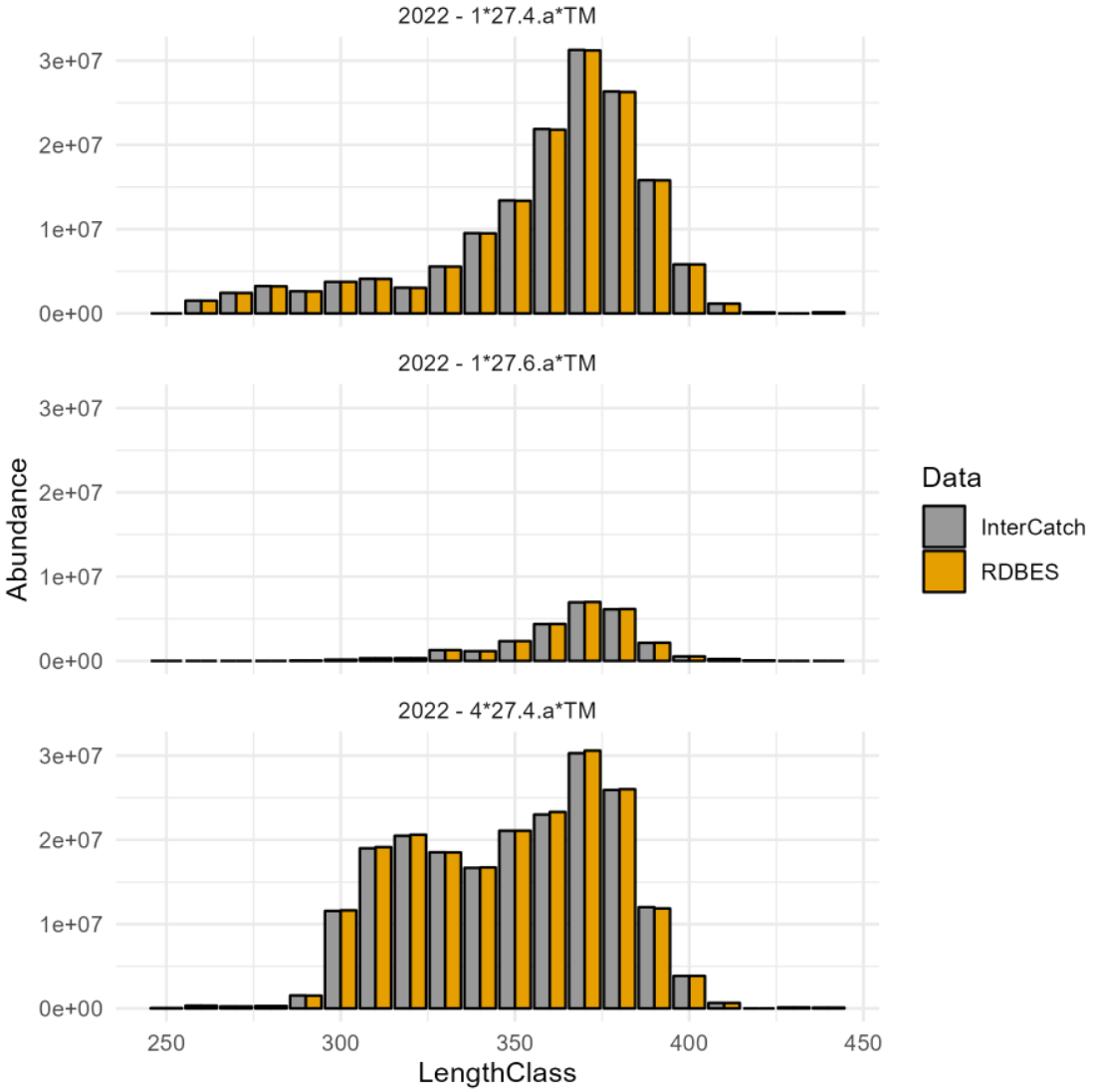


Figure 6 2022 Scottish estimates of abundance at length of mackerel, by domain, as estimated for InterCatch (grey) versus estimating using RDBES data in the TAF repo 2023_all_RDBES_GB-SCT (yellow).

icesConnect didn't work in Raise & TAF Flow, we think this is due to security applied by our institute. This is still to be further tested.

Comparisons of summaries of RDBES CE effort upload and MIXFISH effort files is in progress.

3 Reproduction of stock coordination

It was suggested to start writing the R scripts from the current stock extracts that can be downloaded from InterCatch. Those files show all the imported data related to the relevant stock (stock overview, numbers at age/length, mean weights at age/length). In a continuation of this work, the aim should be to use the output from ToR a as an input. The progress towards reproducing stock estimates is summarised below.

Table 2 Progress on stock coordination reproduction

Participants	Prior effort	Stocks	Quantity	TAF progress	Reproduction
Belgium	'None'; "R&T_22"	Celtic Sea sole (sol.27.7fg)	*CANUM_LAN, WECA_LAN, CANUM_DIS, WECA_DIS, CATON_DIS*	*Partial*	*Reproduced (outcome of R&T_22)*
Belgium	'None'; "R&T_22"	North Sea brill (bll.27.3a47de)	*CANUM_LAN, WECA_LAN, CANUM_DIS, WECA_DIS, CATON_DIS*	*Partial*	*Reproduced/Partial?*
Denmark	None	Sprat 3a and 4 (sprat 3a4)	*CANUM_LAN-WECA_LAN*	Started	
Estonia	None	Baltic sprat (spr.27.22-32)	*CANUM_LAN*	Did not start	
Ireland	none	had.27.7b-k	CANUM_CATCH-WECA_CATCH	None	Qualitatively Reproduced
Latvia	None	Gulf of Riga Herring (her.27.28)	*CANUM_LAN-WECA_LAN*	Started	Started
Netherlands		North Sea plaice (ple.27.420)	*CANUM_LAN, WECA_LAN, CANUM_DIS, WECA_DIS, CATON_DIS*	started	started
Norway	2 weeks	North Sea saithe (pok.27.3a46)	All CATON, CANUM and WECA (including IC postprocessing)	Complete	Reproduced (but based on IC export format).
UK (England)	Days	Northern shelf cod (cod.27.46a7d20)		Partial	Started

Prior effort: If any work was done in adapting the estimate(s) to RDBES and TAF before the workshop. Use categories: 'None', 'days', 'weeks''. The text "R&T_22" indicates that previous work was done during the Raise&TAF workshop in 2022.

Stocks: stocks reproduced, uses ICES stock codes

Quantity: Which quantity was estimated.

- *CANUM_LAN*: numbers-at-age or length for the landings
- *WECA_LAN*: mean weights-at-age or length for the landings
- *CANUM_DIS*: numbers-at-age or length for the discards
- *WECA_DIS*: mean weights-at-age or length for the discards
- *CANUM_CATCH*: numbers-at-age or length for the catch
- *WECA_CATCH*: mean weights-at-age or length for the catch
- *CATON_DIS*: total discards in weight

TAF progress: To what extent TAF was used in the reproduction.

- *Started*: Did not organise code according to TAF standards.
- *Partial*: the estimate was organised with the standard TAF directories and scripts (icesTAF::taf.skeleton)
- *Complete*: the estimate can be re-run with icesTAF::sourceAll

Reproduction (started from the InterCatch input files):To what extent the reproduction was successful:

- *Started:* Did not progress far enough to conclude on reproduction.
- *Reproduced*: Results are reproduced to the satisfaction of the person implementing the reproduction.
- Partial: Results are not reproduced to the satisfaction of the person implementing the reproduction, but the reasons why are thought to be understood and described in this report.
- *Issues*: Results are not reproduced to the satisfaction of the person implementing the reproduction,

3.1 Report from reproduction studies

Belgium

During this workshop, Belgium worked on two stocks: sol.27.7fg and bll.27.3a47de. In the Raise&TAF workshop in 2022, we managed to replicate the InterCatch raising and allocation procedures, including the aggregation step, with small deviations related to rounding (focussing on sol.27.7fg and sol.27.7d). During this workshop, we focussed on getting this code in the TAF structure and incorporating automation/generalisation steps. We made good progress for both stocks.

- **Sol.27.7fg** (https://github.com/ices-taf/2023_sol.27.7fg_RDBES_combined):
  - data.R: Awaiting the file format, which will be produced from the national estimations and considering the current transitional period, the data.R script reads the InterCatch input files from the bootstrap/initial/data folder. Those input files (csv files with the InterCatch Exchange Format) are then converted to the InterCatch stock extracts (StockOverview.txt, NumbersAtAgeLength.txt and MeanWeightAtAgeLength.txt) and copied in the /data folder.
  - method.R: The method.R script sources 3 more method scripts. The first one (method_1_Group_definition.R) calculates a number of datasets, which are needed to define the different groups in which the strata will be clustered for discard raising and age allocation. Method_2_Discard_raising.R contains

automated code to perform the discard raising according to the groups defined in method_1. Method_3_allocation.R will contain the code for the age allocation (not adopted yet).
  - output.R: This script will contain the aggregation step (part of the export step in InterCatch) (not adopted yet).
  - report.R: The plan is to generate a markdown containing a report of the previous steps (not adopted yet).
- **Bll.27.3a47de** (https://github.com/ices-taf/2023_bll.27.3a47de_RDBES_combined):
  - data.R: Considering the obscurity of the file format, which will be produced from the national estimations, the data.R script copies the StockOverview.txt, NumbersAtAgeLength.txt and MeanWeightAtAgeLength.txt from the bootstrap/initial/data folder in the /data folder. This is a provisional solution.
  - method.R: The method.R script sources 3 more method scripts. The first one (method_1_Group_definition.R) calculates a number of datasets, which are needed to define the different groups in which the strata will be clustered for discard raising and length allocation. Method_2_Discard_raising.R contains automated code to perform the discard raising for brill according to the stock annex. Method_3_allocation.R performs the length allocation.
  - output.R: This script contains the aggregation step (part of the export step in InterCatch). However, this part is not adjusted to the TAF structure yet and does not include automation/generalisation of code.
  - report.R: This script is not adapted to the brill stock yet. The plan is to generate a markdown containing a report of the previous steps.

In conclusion, the stock assessor estimation was reproduced, but not all parts are in TAF format yet. Additional work will need to be carried out in a future workshop.

Denmark

The stock coordination for spr.27.3a4 is a regional estimation done with detailed sample and aggregated, but detailed, landings data. Data is submitted by MS in a very specific format and the estimation is presently done with SAS scripts. In the past it has been checked that the RDBES data model has the information needed for the estimation, plus much more. InterCatch has never been used in the process.

The aim for this WK in relation to spr.27.3a4 was not to migrate the SAS script to R and reproduce the estimation, that will be for the benchmark in 2024, but to make sure that TAF and the exchange of detailed data is supported by the RDBES.

The TAF structure supports the present estimation workflow and the suggestion about roles and permissions from this WK seems to allow the spr.27.3a4 SC access to detailed data. In this WK it was suggested that the exchange of detailed data should be an RDBESDataObject pushed from a national repository to an intermediate database. This seems to be a good idea, since that allows for documented national processing of data before submitting and also puts the responsibility for the quality of the submitted data on a national data submitter. Detailed data could also come directly from the RDBES database, but then the direct involvement of the national submitter would be lost.

(The above is also true for san.27.1r-san.27.6)

Estonia

We used the Nationally submitted InterCatch files only to validate the estimation results of SPR landing of Estonian fleet. Hence no stock coordination data validation was done.

Ireland

had.27.7b-k

For this stock, InterCatch is not used to allocate unsampled discards and landings. Instead this is already done in an R script. This script was reviewed and accepted at the most recent benchmark. Therefore, there is no further work to do under ToR b for this stock.

Netherlands

ple.27.420

The discards raising and age allocation for this stock is more complicated. Many strata are defined (see Table 3) by grouping metier categories, subarea and season. The borrowing of strata for imputation following a procedure of first searching for other subarea, then other quarter, and last other similar metier. Additionally, some extra decisions were made to exclude "strange" discards rate from specific country due to a specific fisheries or lack of sample size.

In the workshop, we focus to develop a generic script to identify groups strata as shown in table and identify the matched strata to impute, following a subarea-season-metier borrowing procedure. Additionally, we focused on visualize the available strata information (e.g. discards rate) with sample size. This would help the stock coordinator to gain more information about the discards raising procedure.

Table 3 Grouping strategies to raise discards and allocate age structures

Group for discards raising and age allocation*	quarter + area	description
TBB<100(excluding CRU_16-31)	Each quarter + 4/320	Beam trawl, smaller mesh size
TBB>=100	Each quarter + 4/320	Beam trawl, larger mesh size
TBB/OTB_CRU_16-31	Each quarter + all area	shrimper
OTB/OTM-CRU/DEF/SPF<100(excluding CRU_16-31)	Each quarter + all area	Otter trawl, smaller mesh size
OTB/OTM-CRU/DEF/SPF>=100	Each quarter + all area	Otter trawl, larger mesh size
SSC/SDN<100	Each quarter + all area	Seines, smaller mesh size
SSC/SDN>=100	Each quarter + all area	Seines, larger mesh size
GNS/GTS/GTR<100	Each quarter + all area	Gillnet, smaller mesh size
GNS/GTS/GTR>=100	Each quarter + all area	Gillnet, larger mesh size
Others	All quarter + all area	All other metiers

*** all_0_0 are treated as >=100. TBB/OTB_CRU_16-31 is raised from OTB_CRU<100, because several countries have extremely high discard rates and their fisheries might have different regulations.**

Norway, IMR

Pok.27.3a46:

Scripts were formerly available and partially integrated in TAF for reproducing the InterCatch raising for data from 2021, following WKRDBES-Raise&TAF (2022) and WKRDBES-Raise&TAF-Flow (2023).

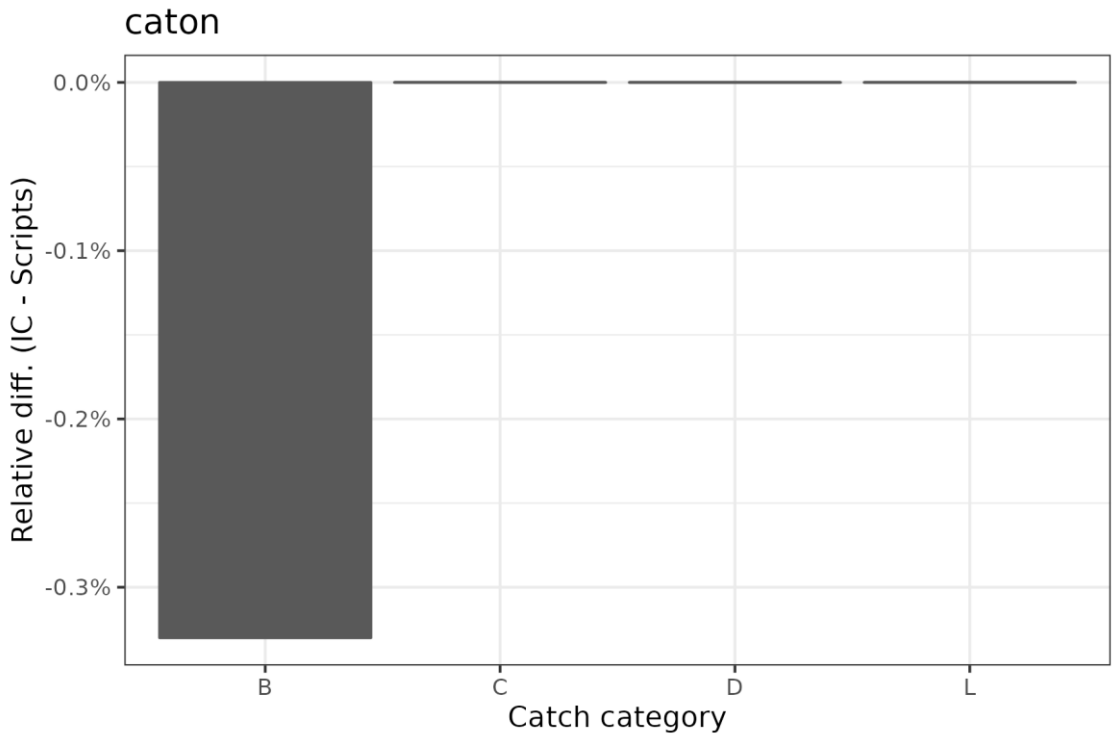
The aim, during the current workshop, was to:

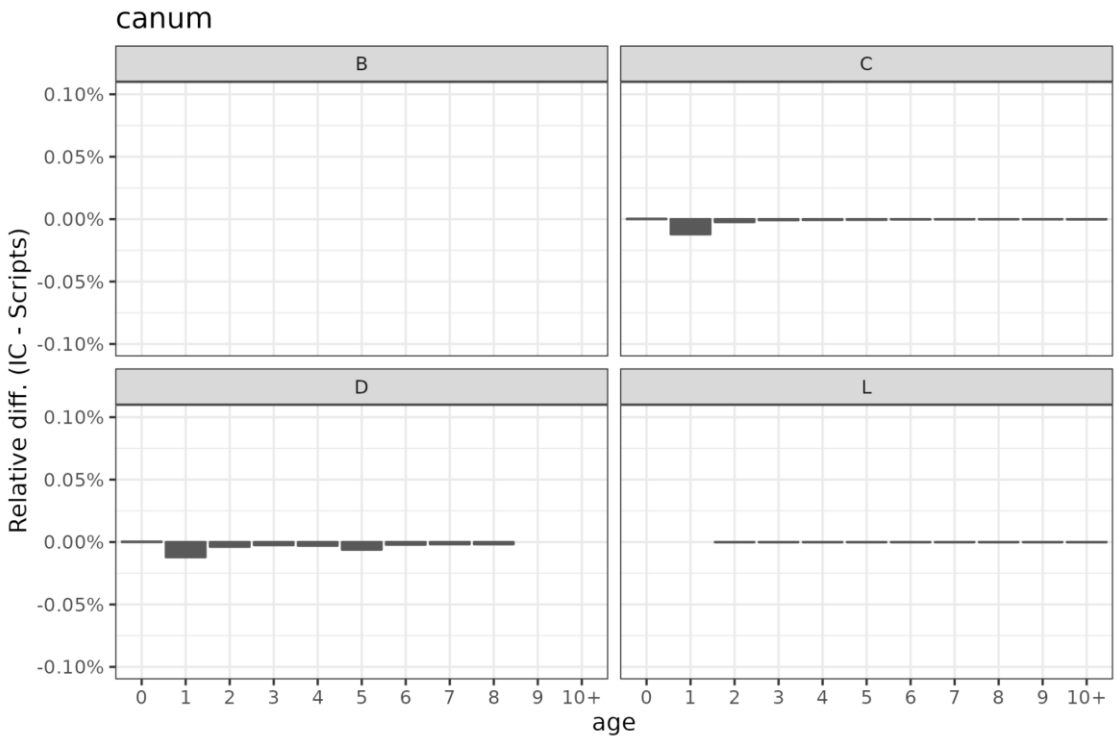
1. adapt the scripts to 2022 data.
2. Complete integration in the TAF workflow, in particular the formatting of the output, including generation of files ready to append to the SAM model inputs (cn.dat, cw.dat, lf.dat, etc.).
3. automatically report on matching performance and validate raising procedures for the 2022 data.
4. validate (and better integrate) the section of the workflow that replaces post-processing of the InterCatch outputs (country-based BMS allocation), by checking outputs against the 2023 stock assessment (SAM) inputs.

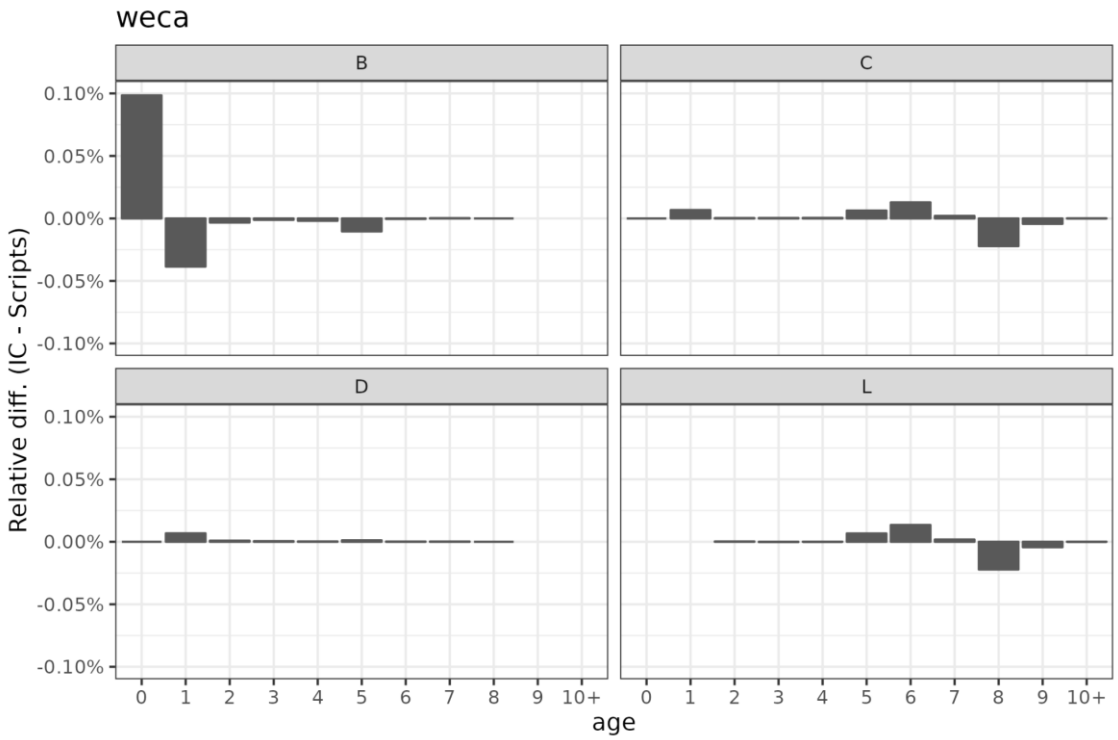
Integration of the workflow in TAF was completed, and adaptation of the scripts to 2022 data carried out successfully (steps 1-2). The definition of groups for imputation of the discard rates, however, remains a bit cumbersome, as it still requires editing two different scripts. It also lacks,

unlike group definitions for age imputations, the flexibility of matching provided discard data to several groups (as it could be convenient when, for instance, most countries provide landings and discards - then raised accordingly - per quarter, while one country needs its discards rates to be estimated for the whole year at once, based on the same quarterly discards from other countries).

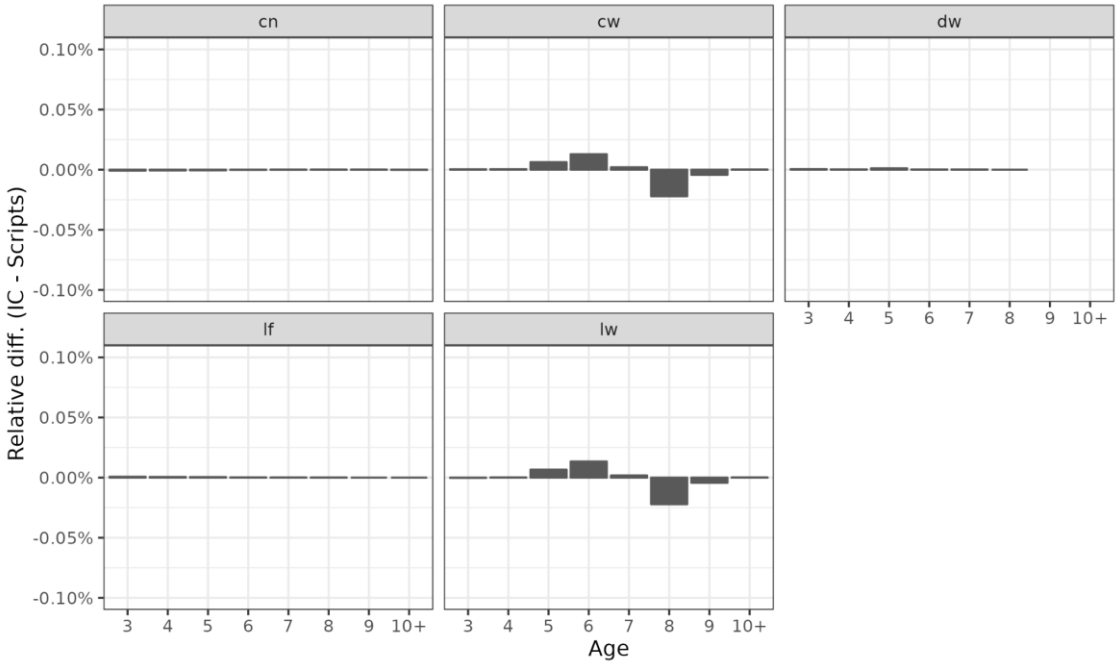
Regarding step 3, the TAF results were very close to the InterCatch raised outputs: with the exception of the overall catch weight for BMS, of which the 0.3% difference can be tracked down to the rounding in the caton.txt file (0.151 vs 0.1515), and weight at age for BMS of the youngest classes (not used in the assessment), all differences were typically within 0.025% (see graphics below for respectively caton, canum and weca). The raising procedures seem therefore consistent.


caton
0.0%
-0.1%
-0.2%
-0.3%
Relative diff. (IC - Scripts)
B
C
D
L
Catch category


canum
B
C
D
L
0.10%
0.05%
0.00%
-0.05%
-0.10%
Relative diff. (IC - Scripts)
0 1 2 3 4 5 6 7 8 9 10+
age



The outputs formatted to match the stock assessment inputs, when compared to data used in the 2023 stock assessment (step 4), revealed a near perfect match, here again typically within 0.025% difference for all SAM inputs usually derived from InterCatch data:



This last step would however benefit from being validated on InterCatch data prior to 2021, when a larger amount of BMS data was declared for NS saithe, in particular originating from Norway (then attributed to landings, unlike other countries for which they are attributed to discards).

4 Roles and permissions

During the workshop the following tasks were undertaken:

1. Further specify the required RDBES/TAF roles and permissions based on the work previously done in the 2022 WKRDBES_Raise&TAF and WGRDBESGOV meetings. Ensure the descriptions are detailed enough for the ICES Secretariat to implement.
2. Agree types of TAF database outputs and metadata.

During the first WKRDBES Raise&Taf workshop a number of roles were identified: Data submitter, National estimator, Stock coordinator, Stock assessor, and WG member. During discussions about these roles, it was recognised that some further refinements were necessary – the current roles proposed are shown below. Note that most role assignments need approval – the proposed hierarchy of approvers is:

1. DCF National Correspondent (if relevant),
2. ACOM Member,
3. ICES Delegate.

(However, it was noted that workshop participants had different views on this order of approvers so it might not be possible to have a single system which is efficient for all countries.)

- **RDBES CS Data Reader**
  - Read access to RDBES CS data by *either*:
    - specified *country codes* / *submitter organisation*
    - *specified stock*
  - An organisation which submitted data automatically has access to that data
    - Any other access needs approval
- **RDBES CE/CL Data Reader**
  - Read access to RDBES CE/CL data for specified *country codes*
  - *Role needs approval*
- **RDBES CS Data Submitter**
  - Same rights as RDBES CS Data Reader plus:
    - Write/delete access to RDBES CS data for specified *country codes* / *submitter organisation*
    - *Role needs approval*
- **RDBES CE/CL Data Submitter**
  - Same rights as RDBES CE/CL Data Reader plus:
    - Write/delete access to RDBES CE/CL data for specified *country codes*
    - *Role needs approval*
- **National Estimator**
  - Read access to RDBES CS data for specified *country and organisation(s)*
  - Read access to RDBES CE/CL data for specified country
  - Read/write access to "**RDBES_Estimation_Input**" and "**RDBES_Estimation_Overview**" outputs for specified *country codes*
  - Read/write access to **RDBES_Estimation_Output** for specified *country codes*

- Role needs approval

- **Regional Estimator**
  - Read access to RDBES CS data by *either*:
    - Specified stock(s)
    - Specified Sampling schemes
  - Read access to required RDBES CE/CL data for:
    - specified countries and stock(s) (just stock area for CE)
    - For records where stock is not known then access can be controlled by:
      - countries, quarter, area, rectangle, species, fisheries management unit
      - **Note**: CL data often needs to pre-processed to correctly identify stocks e.g. grouping species
      - **Note**: The implementation of this is not easy and the proposed solution still needs to be confirmed as feasible by the ICES Secretariat
  - Read/write access to "**RDBES_Estimation_Input**" and "**RDBES_Estimation_Overview**" outputs for specified *country codes and stock*
  - Read/write access to **RDBES_Estimation_Output** for specified *country codes and stock*
  - Role needs approval
- **Restricted Estimator**
  - Read access to RDBES CS data by *country, organisation(s),* stock(s), Sampling scheme(s)
  - Read access to RDBES CE/CL data for specified country and stock (just stock area for CE)
  - Read/write access to "**RDBES_Estimation_Input**" and "**RDBES_Estimation_Overview**" outputs for specified *country codes and stock*
  - Read/write access to **RDBES_Estimation_Output** for specified *country codes and stock*
  - Role needs approval
- **Stock Coordinator**
  - Read access to "**RDBES_Estimation_Overview**" for specified stocks
  - Read access to "**RDBES_Estimation_Input**" for specified stocks and country codes **only if country allows it**
  - Read access to **RDBES_Estimation_Output** for specified stocks
  - Read/write access to **Stock_Coordination_Output** for specified stocks
  - Role needs approval (but this is already handled by the existing ICES Advice processes)
- **Stock Assessor**
  - Read access to "**RDBES_Estimation_Overview**" for specified stocks
  - Read access to "**RDBES_Estimation_Input**" for specified stocks and country codes **only if country allows it**
  - Read access to **RDBES_Estimation_Output** for specified stocks
  - Read access to **Stock_Coordination_Output** for specified stocks
  - Read/write access to **Stock_Assessment_Output** for specified stocks

    - Role needs approval (but this is already handled by the existing ICES Advice processes)
- **WG Member**
    - Read access to "**RDBES_Estimation_Overview**" for specified stocks
    - Read access to **RDBES_Estimation_Output** for specified stocks
    - Read access to **Stock_Coordination_Output** for specified stocks
    - Read access to **Stock_Assessment_Output** for specified stocks
    - Role needs approval (but this is already handled by the existing ICES Advice processes)

Following the update to the roles and permissions the previously proposed types of outputs were also reviewed. The metadata needed in the TAF database was also proposed.

The first outputs concern access to data in the RDBES

- **RDBES CS Data** (called e.g. "2020_DK" in previous reports)
    - CS data submitted/download from RDBES
    - Access permissions controlled by RDBES according to roles as defined above
- **RDBES CE/CL Data** (called e.g. "2020_DK" in previous reports)
    - CE/CL data submitted/download from RDBES
    - Access permissions controlled by RDBES according to roles as defined above

The following outputs are files within TAF repositories – an entry is created for them in the TAF database along with the required metadata. Access to the outputs is controlled by the TAF database using the metadata according to the roles defined above.

- **RDBES_Estimation_Input** (was called "_detail" in previous reports)
    - When data is submitted to stock assessment data calls it is often pre-processed – this can include tasks like grouping or re-assigning metiers, species, and areas. The RDBES_Estimation_Input contains this pre-processed RDBES data that is then used as an input to estimation.
    - Includes CS, CL, CE data
    - The format of this output is an RDBESDataObject (as defined in the RDBEScore package https://github.com/ices-tools-dev/RDBEScore)
    - TAF database metadata required: Year; Stock; Country
- **RDBES_Estimation_Overview** (was called "_overview" in previous reports)
    - This is a report summarising the landings and effort data, and the national stock estimation process.
    - It is planned to be in an HTML format but the content of this report is not defined yet
    - TAF database metadata required: Year; Stock; Country
- **RDBES_Estimation_Output**
    - This will include: landings and effort data for use in stock coordination and assessment, and biological estimates derived from sampling data
    - The format of this output was discussed during the workshop – see Section 5.
    - TAF database metadata required: Year; Stock; Country

- **Stock_Coordination_Output**
  - Outputs from the stock coordination process (e.g. "canum_weca")
  - The format of this output needs to be discussed and agreed.
  - This is used as an input for stock assessment
  - TAF database metadata required: Year; Stock;
- **Stock_Assessment_Output**
  - Retain current format
  - TAF database metadata required: Year; Stock;

In the previous "Workshop on Raising Data using the RDBES and TAF" https://doi.org/10.17895/ices.pub.21995141 report an example of cod.27.21 was used to illustrate the roles and permissions - this stock requires data from 3 countries: Denmark, Sweden and Germany. This example has been updated to illustrate the refined roles and permissions.

- "R" means read-only access
- "R*" means read-only access only if relevant country allows it
- "CRUD" means Create, Read, Update, Delete access

For the example of cod27.21 access to RDBES data for the different roles is shown below:

RDBES data type					
	DK, 2020				
	CS Data Reader	**CE/CL Data Reader**	**CS Data Submitter**	**CE/CL Data Submitter**	**National Estimator**
CS Data (DK, 2020)	R		CRUD		R
CE/CL Data (DK, 2020)		R		CRUD	R
	SE, 2020				
	CS Data Reader	**CE/CL Data Reader**	**CS Data Submitter**	**CE/CL Data Submitter**	**National Estimator**
CS Data (SE, 2020)	R		CRUD		R
CE/CL Data (SE, 2020)		R		CRUD	R
	DE, 2020				
	CS Data Reader	**CE/CL Data Reader**	**CS Data Submitter**	**CE/CL Data Submitter**	**National Estimator**
CS Data (DE, 2020)	R		CRUD		R
CE/CL Data (DE, 2020)		R		CRUD	R

For the example of cod27.21 access to TAF outputs for the different roles is shown below:

Output type	**Role**					
	DK, 2020	**SE, 2020**	**DE, 2020**	**Cod.27.21**	**Cod.27.21**	
	National Estimator	**National Estimator**	**National Estimator**	**Stock Coordinator**	**Stock Assessor**	**WG Member**
Estimation_Input (DK, 2020, Cod.27.21)	CRUD			R*	R*	
Estimation_Overview (DK, 2020, Cod.27.21)	CRUD			R	R	R
Estimation_Output (DK, 2020, Cod.27.21)	CRUD			R	R	R
Estimation_Input (SE, 2020, Cod.27.21)		CRUD		R*	R*	
Estimation_Overview (SE, 2020, Cod.27.21)		CRUD		R	R	R
Estimation_Output (SE, 2020, Cod.27.21)		CRUD		R	R	R
Estimation_Input (DE, 2020, Cod.27.21)			CRUD	R*	R*	
Estimation_Overview (DE, 2020, Cod.27.21)			CRUD	R	R	R
Estimation_Output (DE, 2020, Cod.27.21)			CRUD	R	R	R
Stock_Coordination_Output (2020, Cod.27.21)				CRUD	R	R
Stock_Assessment_Output (2020, Cod.27.21)					CRUD	R

5 Estimation file formats

We formed a subgroup to evaluate a draft proposal for a standard TAF output from “national estimate”-TAF repositories. The draft proposal was prepared by the workshop chairs, as a suggestion for a way to generalise upon the InterCatch Exchange format to form a generic and flexible way to communicate domain estimates from fisheries. The subgroup evaluated the draft and identified some issues. The subgroup prepared a revision of the format draft and noted the rationale for changes made. This document is included as Annex 2:. n addition to a proposal for format, and explanation about the restrictions imposed, it contains an overview of requirements for the format both identified at this workshop and at previous workshops.

The subgroup also set forth to suggest a standard TAF output format for “combined estimate”-TAF repositories (output from stock coordination or stock estimates) but did not make noteworthy progress towards that goal.

6 TAF, GitHub and confidential data

RDBES data needs to be available to the TAF server - currently this is done by pushing the data to a private GitHub repository so that the TAF server can access it. However, some workshop participants said they might not be allowed to push RDBES data to GitHub - this would prevent it being used in TAF.

This workshop considered how confidential RDBES data can be used within TAF without requiring it to be uploaded to GitHub - the potential solutions were identified and evaluated.

Option #	Possible solution	Pros	Cons
1	Use a secure ICES web service to get RDBES data	RDBES data is never stored on GitHub. Easier and more efficient than copying files.	Would need to be developed – currently not funded
2	User manually copies data to TAF server using TAF app	RDBES data is never stored on GitHub.	Manual work required to copy files around. Would need to be developed.
3	Store data on ICES SharePoint and get it using a web service	RDBES data is never stored on GitHub. The function to get files from ICES SharePoint already exists.	Increases complexity and dependencies. Need to mimic RDBES/TAF authentication roles and rules in SharePoint. Not a good long-term solution.
4	Encrypt RDBES data file for storage on GitHub	Relatively easy to implement.	Data is still stored on GitHub (although encrypted) so doesn't solve the problem.
5	ICES could host its own git server (e.g. a basic git server, or GitLab)	RDBES data is never stored on GitHub.	Would need to be installed, maintained, and supported
6	Use an internal ICES function to copy RDBES data to the TAF server.	RDBES data is never stored on GitHub.	Would need to be developed – currently not funded.

Based on this evaluation the workshop believes that the best way forward is for the ICES Secretariat to develop a secure web service to fetch RDBES data. This would have the benefits of: i) being a more efficient way for scripts in TAF to import data (as compared to manually downloading and copying RDBES data files), and ii) remove any potential problems related to storing RDBES data on GitHub.

7 Catch/Effort Overview case study

During the workshop the subgroup "Catch/Effort Overview case study" (SG) aimed to the adaptation and development of code in order to produce data overviews specular with those produced by the European Regional Coordination Group (RCG) Intersessional Subgroup (ISSG) "Regional DataBase (RDB) catch, effort and sampling overviews" but based on the Regional Database and Estimation system (RDBES) format. Furthermore, the subgroup aimed to reach its goal in a fully reproducible manner, i.e. consistently with the ICES Transparent Assessment Framework (TAF).

The SG´s work consisted in different activities, namely: i) structure the architecture of the RDB catch, effort and sampling overviews code in order to fit the TAF, ii) trial of the code supported by RDBES related functions and pre-existing RCG ISSG RDB catch, effort and sampling overviews scripts, iii) reporting the developments.

7.1 Fit of the RDB catch, effort and sampling overviews in TAF

The TAF is an online resource that, besides stock assessment purposes, may also include other types of assessments stemming from ICES advice (*e.g.* fishing impacts, survey indices and fisheries overviews). For this purpose, the SG worked on producing a structure for reporting fisheries overviews consistent with the TAF. The discussions on the workflow to include these reports on TAF resulted in the structure reported in Annex 3:. The main challenge in this activity was fitting the product of the RCG ISSG in the TAF, which allows the production of detailed reports as one of its output. In order to fit the overviews into the TAF, the subgroup agreed on:

- The folder utilities_files to gather:
  - A folder named utilities_files/settings, containing a script that allows to govern the overall R settings.
  - A folder named utilities_functions, holding the functions supporting the production of the overviews.
  - A folder named utilities_graphical_parameters, to gather the graphical parameters supporting the creation of plots for the overviews.
- The folder bootstrap/initial/data to hold the data needed to perform the conversion of RDBES in RDB and thus to produce the overviews. This folder contains data processed by several data preparation functions together with their outputs. These are later used by the scripts producing the overview.
- The folder data containing the scripts sourcing the code needed to obtain the set of data to be processed and the script to process them.
- The folder report to gather the files supporting the execution of the code of the report (or holding the information that appears there-in).
- The report RMarkdown file, in turn, to produce the overview as the ultimate result of the outlined code structure.

Secondly, each section of the produced structure was filled with files contributing to the production of the overviews. These files mainly consisted in R scripts: adaptations of already existing scripts together with code developed *ad hoc*. In fact, to maximise the efficiency of the workflow of the SG participants and test the adaptation of the RDB catch, effort and sampling overviews to the TAF, the SG agreed to start with the conversion of the RDBES format in the RDB format.

In detail, commercial landing (CL) and commercial effort (CE) tables were converted from RDBES format to RDB format. The conversion was based both on code developed *ad – hoc*, pre-existing script (the SG is grateful to Nuno Prista, Inst. Marine Research, SLU Aqua, Sweden for the provision of several useful functions). The main challenges in this case concerned the extraction of information from the different tables contributing to the RDBES format and smoothing out the differences in the formats (e.g. column names differing between the RDB and RDBES format).

7.2 Run of the overviews code

The resulting data, converted from RDBES to RDB, were then processed via a pre-existing script aimed at the production of the input data needed to produce the overviews. The latter was in turn used to feed the code producing the RDB catch, effort and sampling overviews via an RMarkdown document. Both the latter scripts were developed at the RCG ISSG RDB catch, effort and sampling overviews and were re-adapted when needed. Since the RCG ISSG "RDB catch, effort and sampling overviews" has multiple complex outputs and time was a limiting factor, the SG agreed on targeting the RDB annual fisheries overview for this exercise. Among the others RCG ISSG "RDB catch, effort and sampling overviews" outputs this is the less complex overview as it reports data belonging to a single year. This was found to be a timeframe consistent with the example RDBES data provided to the SG in the ICES WKRDBESRaise&TAF2 context.

Once the scripts are articulated in the TAF structure, the connection of the codes is performed using scripts to source them. The SG aims to perform a series of runs of the connected code in order to test the production of the RDB catch, effort and sampling overviews using RDBES data in TAF.

7.3 Reporting

Along the SG work, the participants updated the code adaptations and developments on the ICES TAF GitHub repository "2023_CatchEffort_RDBES_Combined" (https://github.com/ices-taf/2023_CatchEffort_RDBES_Combined/tree/main).

The conversion of the code from RDBES to RDB was successful both for CE and CL. This constitutes an important accomplishment as it was set as one of the goals for 2023 – 2024 in the latter European RCG ISSGs meeting (RCG NANS&EA RCG Baltic 2023, ref. in 7.4). The code for conversion is going to be available at the mentioned ICES TAF repository making the conversion of the tables fully reproducible, given the availability of the data. The data – preparation script is run to prepare the input data needed to produce the type of information to be reported in the overviews. The code for producing the annual overviews can then be tested and run with this RDBES converted data to evaluate the results obtained. At this moment, the RDB overviews are produced by region but one of the aims for the future will be to evaluate the possibility of having a generalised, instead of region specific, code for the overviews in TAF.

In this work we initiated the process of integration of the overviews, based on regional databases, in the TAF. Furthermore, we adapted the code already developed at RCG ISSG level to the new RDBES format.

In order to place a second step towards the development of overviews fully built on the RDBES format, we suggest that the further developments should focus on replacing the conversion of the code with scripts treating the RDBES format directly. We see the work developed here as complementary to the one developed at the RCG ISSG RDB catch, effort and sampling overviews. We thus forecast that the work developed so far will support the roadmap of the RCG ISSG and the work developed in that context will allow for improvements in the code and its

structure in the ICES TAF in the near future. We recognize the need for further work to be carried on in order to obtain fully reproducible RDBES overviews in TAF. For this reason, we suggest this exercise to be deepened and expanded further in the next ICES WKRaise&TAF.

7.4 References

RCG NANS&EA RCG Baltic 2023. Regional Coordination Group North Atlantic, North Sea & Eastern Arctic and Regional Coordination Group Baltic. 2023. Part I Report, 79 pgs. Part II Decisions and Recommendations, 13 pgs. Part III, Intersessional Subgroup (ISSG) 2021-2022 Reports, 320 pgs. (https://datacollection.jrc.ec.europa.eu/docs/rcg)

8 RDBES/TAF issues identified

Technical issues with TAF database services

The TAF database will be used to flag files as being an output from an estimation repository, such that they can then be used as an input to a different repository. During the workshop this facility was not working so could not be tested. This is a key tool that will be used to link RDBES/TAF repositories together and ensure data can flow from the RDBES through the advice process. Until this system is working reliably it will not be possible to fully test the proposed RDBES/TAF workflow.

Technical issues with RDBES

During the workshop, a number of participants could not download data from the RDBES that they had previously uploaded. This was due to the data being mistakenly deleted by a software bug. The bug has now been fixed and countries affected were asked to re-upload their data.

Upload Logs

RDBES has developed strict data definitions, and the RDBES data call invites high quality data ready for use in ICES estimates. Data submitters may, in this context, be wary about making approximations that they may find later on to have been interpreted strictly. At the same time, proper testing might require partial data, or data to be provided with narrow assumptions about usage. With this in mind, data submitters have been encouraged to submit preliminary data if it was not possible to submit complete data (this was also done for the Raise&TAF workshops). In order for the RDBES to be fully operational the quality of the data needs to be clear and should not be left to assumptions. The strict data requirement may be a hindrance to testing and gradual adaptation. Unless a testing environment can be provided for both data submitters and data users parallel to the production environment, some mechanism must be implemented to make sure data users can be informed about caveats. During the WKRDBES-POP workshops, some examples of upload logs modelled on the RDB upload logs were evaluated. These could provide a means for compromising between strict definitions and general use on the one hand and need for testing and gradual adaptation on the other.

RDBES/TAF Repository Structure

It has previously been agreed that we should have a single RDBES/TAF national estimation repository for each country/year - there are two main benefits to this:

1. It is clear where national estimation scripts and results are found.
2. It avoids a proliferation of RDBES/TAF repositories and the resulting administrative overhead in creating and maintaining them.

During the workshop some participants pointed out that they will have a lot of scripts that are used during estimation and it might be hard to manage and keep track of them within a single repository. A solution to this problem is that it is possible to use folders within a repository to organise scripts. In this report there are examples presented in the Section 2.1 and Annex 3:.

TAF - version control of estimation outputs

Different assessment working groups have different deadlines for their data calls - in between these deadlines the data in national databases can be modified/updated. These types of updates might include:

- Updates to national landing/effort statistics
- Ageing data is appended. National ageing deadlines often mean not all fish will be aged for the earliest data call deadlines, only the species that are actually required for the data call. As the year progresses more fish will be aged and will need to be included in later estimations.

In order to use the most recent data in the advice process these data updates should be submitted to the RDBES.

Running the TAF function taf.boot() will use the data.bib file to create data that can be used in the TAF analysis - in the future it is recommended that this will call a web service to download the latest data from the RDBES.

Running the TAF function icesTAF::sourceAll() re-runs all the scripts to create the "data", "method", "output", and "report" folders. Since the current plan is to have a single national estimation repository for each country the combination of these two functions will create all national estimation outputs using the latest RDBES data. If the RDBES data has been updated, then it might be the case that national estimation output files that have already been used in earlier data calls get re-created with different values (however note that the data in the stock estimation repositories will not change). On a national level this will make it hard to keep track of which version of the file was actually used in a data call.

9 Future reproduction workshops

The Raise&TAF workshops (2022 and 2023) have served several purposes for the RDBES development. They have:

- evaluated the use of TAF as a framework for estimation for RDBES
- encouraged adaptation of RDBES and TAF by data providing institutions
- served as a training arena for the ICES community to get acquainted with RDBES and TAF
- developed suggestions for standardised TAF output formats
- elucidated roles and data access requirements for current and envisioned estimation workflows.
- demonstrated that satisfactory reproductions of submissions to InterCatch can be made based on RDBES data for selected stocks and countries.
- produced potentially re-usable code for tasks shared between roles in the RDBES/TAF system. Particularly for stock coordination tasks.

Their main purpose has been to make an entire reproduction of stock coordination output from detailed national data submitted to RDBES but this has yet to be delivered. However, necessary components of such estimates have been produced. Several countries have been able to satisfactorily reproduce their submissions to InterCatch and have implemented their calculations in TAF, and at least one stock coordination task has been completely implemented in TAF (albeit based on files prepared for InterCatch, rather than actual output from upstream TAF repositories). As the main purpose of the workshop has yet to be completely achieved, we suggest that a workshop with this purpose is arranged also in 2024.

Since the individual components of reproduction have been achieved we think such a workshop could be set up to actually deliver input to ICES stock assessment calculations in 2024 and start transitioning into operational use of the new infrastructure. The workshop will not necessarily have to use all features of RDBES/TAF but aim to incorporate as many as possible while still delivering data consistent with earlier deliveries. For instance, TAF may be used for stock coordination, and even for some national estimates, without it necessarily being used for all national estimates. Or the TAF-project structure could be used without putting it on GitHub (if technical issues or data confidentiality concerns prohibit that). The national workflows that produce data for InterCatch could be put into a TAF structure if results have not been reproduced with RDBES yet, or RDBES data could be used selectively (e.g. using the RDBES CL table for landing statistics) while still relying on national data formats for estimation.

As such a workshop would deliver the first ICES estimates of commercial harvest from RDBES and TAF, it will form an example that later adapters will look to for implementation. In order to avoid temporary solutions to become cemented, it is strongly advisable that technical hurdles identified in this workshop are addressed (see Section 6 and Section 8 and the corresponding recommendations). For the same reason, the work on defining standard TAF outputs from both "national estimates" TAF-repositories and "combined estimate" TAF-repositories should be completed first. That can possibly be achieved if an intersessional group is set up to finish that work.

It would also be advantageous that other tasks addressed by the Raise&TAF workshops are followed up through other channels. We find that the work of evaluating the suitability of RDBES/TAF has progressed to the point that it no longer needs dedicated workshops. The same

consideration applies to the development of roles and data access requirements. The components of the Raise&TAF program that address training needs can possibly be taken over by the TAF training workshops, and perhaps some of these can be dedicated to applications of TAF in the preparation of input data to stock assessment. As argued above, standard output formats should ideally be resolved before official estimates are to be produced.

While the Raise&TAF workshops have been used to develop R-code to perform estimation, this code has mostly materialised as code specialised to specific countries and specific stock. Some code-sharing has been reported between stocks, which is to be expected as estimates at that point are presented in homogeneous, although not yet standardised, form. More generic tools for "national estimate" TAF-repositories have been developed through the WGRDBES-EST, but a similar forum does not exist for stock coordination tasks. In addition to the estimation task itself, visualisation and other decision-making support is also needed for stock coordination. These tasks have until now received little attention in reproduction workshops, but as the workflow for some stocks are maturing, the time is also ripe to look for solutions for standardised software for estimation. For instance, community driven development can be set up as it has been for "national estimate" TAF-repositories through WGRBDES-EST.

10 Conclusion

The WKRDBES Raise&TAF2 reproduced both national estimates and stock coordination estimates for several stocks. For both kinds of estimates some examples were also implemented in TAF repositories. These two tasks were however not carried to completion for the same stocks, and technical solutions for communicating data between TAF projects were not yet available at the workshop. So the workshop was not able to demonstrate a complete data flow that uses TAF to perform all computations from detailed RDBES data to stock coordination results. In addition, the workshop identified some important issues that need to be resolved for RDBES/TAF to be an acceptable solution for several of the participating countries and institutions.

We consider that some of the issues identified at WKRDBES_Raise&TAF2 are blocking general adaptation of RDBES and TAF for commercial harvest estimates. We make the following recommendations to address these:

1. In order to remove the need to store RDBES data in GitHub repositories we recommend that the ICES Secretariat develop a secure web service to fetch RDBES data. This would have the benefits of: i) being a more efficient way for scripts in TAF to import data (as compared to manually downloading and copying RDBES data files), and ii) remove any potential problems related to storing RDBES data on GitHub. (See background and discussion about this topic in Section 6 and Section 8 the WKRDBES_Raise&TAF2 report.)

2. In order to facilitate testing and gradual adaptation to RDBES, WKRDBES_Raise&TAF2 recommends that WGRDBESGOV implement upload logs for the RDBES, modelled on the RDB upload logs. Upload logs should be mandatory with data submission and should be made easily available for all data users. These should facilitate clear communication about any data set that is not completely uploaded, or any data set where approximations have been made that may narrow the use of the data to particular tasks. (See the discussion about this topic in Section 8 of the WKRDBES_Raise&TAF2 report.)

Some other useful lessons were learned at WKRDBES_Raise&TAF2. These are not considered to be blocking general adaptation of RDBES and TAF, but we hope they will be useful in the continued development and would like to bring them to the attention of the WGRDBESGOV 2023.

- The workshop would like to bring to the attention of the WGRDBESGOV and WGTAFGOV the discussion in Section 9 about how future reproduction workshops similar to Raise&TAF may be scoped. In particular, we suggest a workshop inserting the RDBES/TAF into the production of official estimates for at least one stock in 2024, to develop standard output formats through intersessional work, and to arrange TAF-training workshops for data submitters.
- As RDBES/TAF matures, standard software for estimation becomes increasingly useful. Particularly for the stock coordination task, we would suggest that WGRDBESGOV considers ways to initiate such development. See the discussion in in Section 9.

Annex 1: List of participants

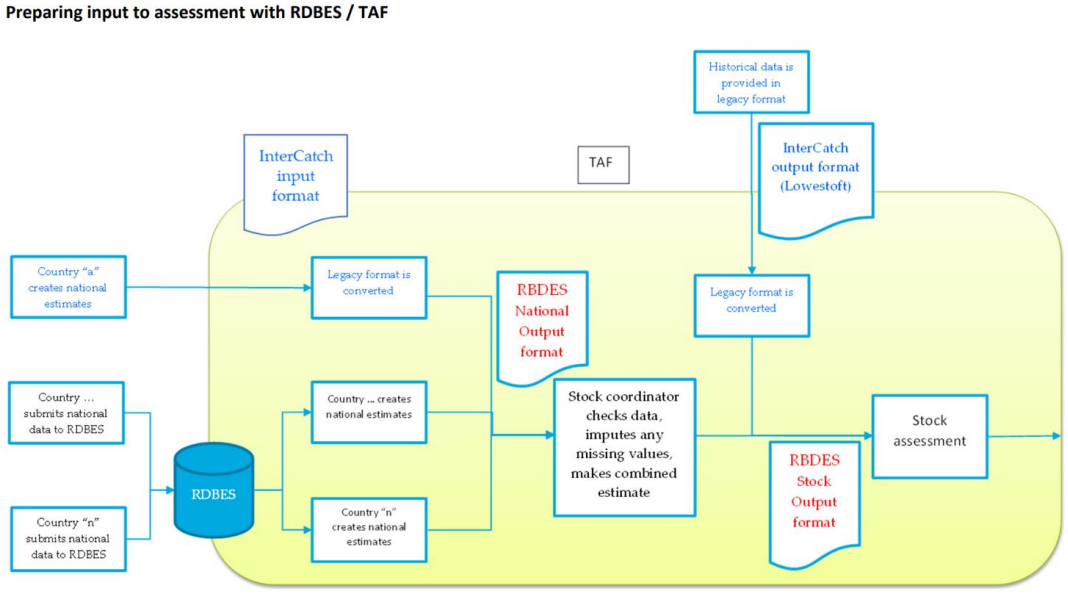
Name	Institute	Country (of institute)	Email
David Currie (Chair)	Marine Institute	Ireland	david.currie@marine.ie
Edvin Fuglebakk (Chair)	IMR	Norway	Edvin.Fuglebakk@hi.no
Afra Egan	Marine Institute	Ireland	afra.egan@marine.ie
Alex Holdgate	CEFAS	UK (England)	alex.holdgate@cefas.gov.uk
Ana Cláudia Fernandes	IPMA	Portugal	acfernandes@ipma.pt
Ana Ribeiro Santos	CEFAS	UK (England)	ana.ribeirosantos@cefas.co.uk
Ane Iriondo	AZTI	Spain	airiondo@azti.es
Anthony Hennessey	IMR	Norway	Anthony.Hennessey@hi.no
Bart Vanelslander	ILVO	Belgium	bart.vanelslander@ilvo.vlaanderen.be
Chun Chen	WUR	Netherlands	chun.chen@wur.nl
Colin Millar	ICES		colin.millar@ices.dk
Cormac Nolan	Marine Institute	Ireland	cormac.nolan@marine.ie
Eros Quesada	SLU	Sweden	eros.quesada@slu.se
Fanyan Zeng	Marine Scotland Science	UK (Scotland)	fanyan.zeng@gov.scot
Hans Gerritsen	Marine Institute	Ireland	hans.gerritsen@marine.ie
Henrik Kjems-Nielsen	ICES		henrikkn@ices.dk
Ivars Putnis	BIOR	Latvia	Ivars.Putnis@bior.lv
Jessica Craig	Marine Scotland Science	UK (Scotland)	Jessica.Craig@gov.scot
Jonathan Rault	Ifremer	France	jonathan.rault@ifremer.fr
Jonathan Stounberg	DTU Aqua	Denmark	jostou@aqua.dtu.dk
Josefina Teruel	IEO-SCIC	Spain	josefina.teruel@ieo.csic.es
Julia Wischnewski	Thünen	Germany	julia.wischnewski@thuenen.de
Karolina Molla Gazi	WUR	Netherlands	karolina.mollagazi@wur.nl
Katja Norén	SLU	Sweden	katja.noren@slu.se
Kirsten Birch Håkansson	DTU Aqua	Denmark	kih@aqua.dtu.dk
Lies Vansteenbrugge	ILVO	Belgium	lies.vansteenbrugge@ilvo.vlaanderen.be

Name	Institute	Country (of institute)	Email
Liz Clarke	Marine Scotland Science	UK (Scotland)	liz.clarke@gov.scot
Lucia Zarauz	AZTI	Spain	lzarauz@azti.es
Maciej Adamowicz	National Marine Fisheries Research Institute	Poland	madamowicz@mir.gdynia.pl
Maria Korta	AZTI	Spain	mkorta@azti.es
María Pan	IEO-SCIC	Spain	maria.pan@ieo.csic.es
Maria Ching Villanueva	Ifremer	France	Ching.Villanueva@ifremer.fr
Nicolas Goñi	LUKE	Finland	nicolas.goni@luke.fi
Nuno Prista	SLU	Sweden	nuno.prista@slu.se
Richard Meitern	Tartu University	Estonia	richard.meitern@ut.ee
Roxanne Duncan	Marine Institute	Ireland	roxanne.duncan@marine.ie
Sara-Jane Moore	Marine Institute	Ireland	sara-jane.moore@marine.ie
Siobhán Moran	Marine Institute	Ireland	siobhan.moran@marine.ie
Sofia Carlshamre	SLU	Sweden	sofia.carlshamre@slu.se
Sofie Nimmegeers	ILVO	Belgium	sofie.nimmegeers@ilvo.vlaanderen.be
Sólva Eliasen	Havstovan, Faroe Marine Research Institute	Faroe Islands	Solvae@hav.fo
Stefanie Haase	Thünen Institute of Baltic Sea Fisheries	Germany	stefanie.haase@thuenen.de
Sven Stötera	Thünen	Germany	sven.stoetera@thuenen.de
Thomas Barreau	MNHN (museum of natural history)	France	thomas.barreau@mnhn.fr
Yves Reecht	IMR	Norway	Yves.Reecht@hi.no

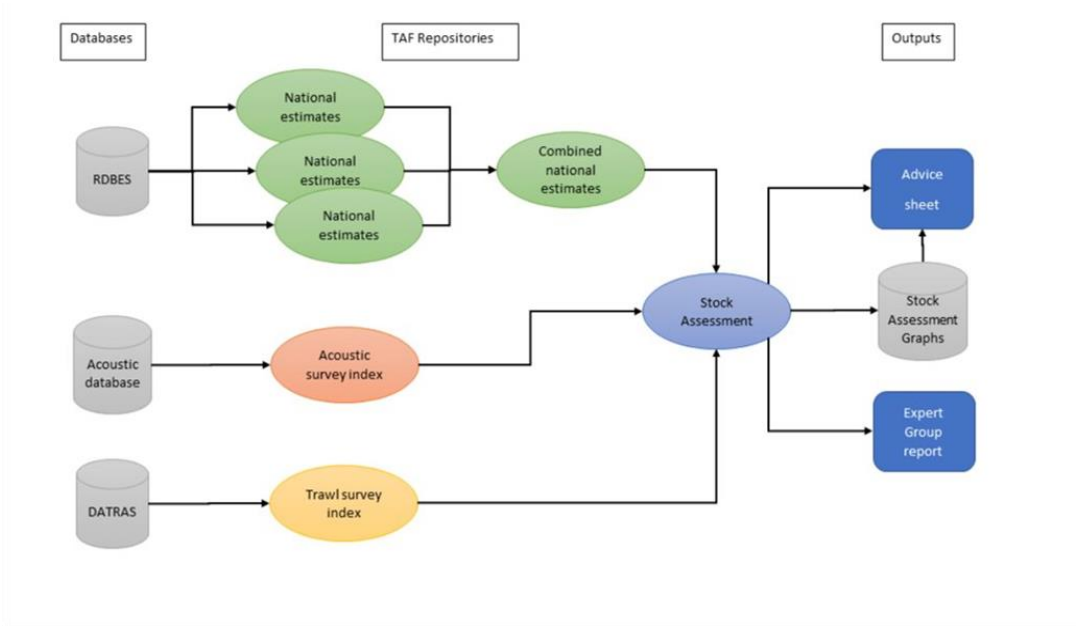
Annex 2: Format draft for RCEF: Regional Catch Estimates Format

An exchange format for national estimates need to be defined in order to standardize exchange, so that general stock-coordination scripts can be developed, and so that the exchange of information can be easily specified in sufficient detail for countries to sign off on data sharing.

The need for this exchange format was reported in the WGRDBESGOV report from 2021 and was summarized there in the figure below:



Since then, RDBES/TAF development has been gravitating towards organising this workflow in two kinds of TAF repositories, one for "National Estimates" and one for "Combined national estimates" (stock coordination):



So the exchange format we require can be considered a standard TAF output type for "national estimates" repositories.

While we are currently working with workflows that provide estimates for each nation, RDBES development has envisioned generalizing this tasks to regional estimation where appropriate. We will therefore suggest to call the exchange format Regional Catch Estimates Format, noting that a nation is just a special case of a region, and we will use the following abbreviations:

RCEF: Regional Catch Estimates Format, referred to in 2021 WGRDBESGOV report as RDBES National Output format. An TAF output type.

ICEF: InterCatch Exchange Format, referred to in 2021 WGRDBESGOV report as Legacy format.

Requirements

Requirements for the format has been developed at earlier workshops:

1. **The exchange format should be compatible with the InterCatch exchange format.** (Raise&TAF 2022)

Users are expected to adapt to RDBES-TAF gradually, and perhaps at different pace for different nations, and since we cannot expect historical data to be generally possible to provide to RDBES. it is desirable to be able to leverage workflows adapted to InterCatch and to leverage historical data archived in InterCatch. It should therefore be possible to convert ICEF to RCEF without loss of information. In order to quickly leverage any downstream software that is adapted to the ICEF format it is also desirable to be able to convert RCEF to ICEF with possible loss of information. The conversion from RCEF to ICEF can probably not be formally guaranteed for all valid RCEF data, without contradicting requirement 2.

2. **Format should be flexible, so that granularity in information can be decided for each use case.** (Raise&TAF 2022)

Higher precision estimates may be available for more aggregated data, and privacy concerns may be easier accommodated by more aggregated data. On the other hand, assessment models may leverage information at higher resolution when data is less aggregated. A flexible format

will facilitate the level of aggregation provided to be agreed upon based on the available precision of estimates and the needs of the assessment model.

Likewise, high precision estimates may be easily obtainable for parts of a fishery, while parts must be inferred by imputation or other techniques. A flexible format will allow more details to be conveyed where the precision is high, and less when speculative inference techniques are required (e.g. imputation).

3. **RCEF should have fields to convey uncertainty in estimates and quality of inference** (Raise&TAF 2022).

These should be suitable for further processing. Particularly it should be possible to infer the variances for sums and weighted averages of domain estimates.

In preparing the Raise&TAF 2023 workshop, another requirement was identified:

4. **The format should be extendable and clearly versioned.**

Stock coordination scripts and other software will be developed before and parallel to the adaptation of users to RDBES-TAF. We also expect revisions of workflows to leverage the information available in RDBEs that was not available in earlier assessment support systems. The format must therefore likely be extended several times, and the format should provide some rules for software developers to ensure backwards compatibility when new format versions are released.

Another requirement suggested by a stock assessor during the Raise&TAF 2023 workshop:

5. **The format should support finer domain definitions in the same file as coarser domains,**

as long as there is no overlap between domains in the same file. E.g., Some domains may be specified by Area and month, and some by Area and Quarter, but all domains specified by month are from different quarters the those specified by Quarter. Similarly, some domains may be specified by SubArea, and others by Division or SubDivision. The latter example may apply within a specific country for the same catch category.

A requirement suggested in plenary discussions at the Raise&TAF 2023 workshop:

6. **The format should facilitate sum-of-product checks of the kinds performed in InterCatch today.**

If domains can be combined into clearly identified domains of the census data, one can estimate these known total landings by the sum of the product of mean weight and total number of fish in each domain (incorporating variables such as age-group or length group into the domain definition).

Draft proposal for RCEF

The product of a national or regional estimate is some statistic reported for a well-defined population that is partitioned into domains, and some aggregation of census data similarly partitioned into domains. We would like to define a format that allows great flexibility in defining such domains for all kinds of estimates that can be produced from RDBES data but remains compatible with InterCatch. So we want to generalize the ICEF into a generic format for conveying domain estimates. The domain of estimates is often referred in terms of the domains that partition the population of catches by the activity leading to catch (e.g. time and place for catch, and the gear used to obtain catch). From the point of view of estimation, also the discrete variables used to partition the population of catches based on biological parameters (age, length-groups, etc.) can be considered part of the domain. We refer to this as *activity domains* and *fish domains*, respectively. We will refer to the quantities reported for each domain as *value fields*.

If we want to retain backwards compatibility with InterCatch. We need to make sure we can provide stock assessors with the data that was provided through InterCatch. We can use the format definition to check if we can provide all that InterCatch could provide. If there are unsupported cases, we can possibly go back and check if that data has ever been provided.

We can model the types of regional outputs we need from InterCatch, and define three kinds of tables corresponding to InterCatch HI, SI and SD records:

- EFFORT: Total effort for a nations fleet or some partition of it, corresponding to HI records. Ideally this can be based on field definitions provided for the CE table. EFFORT may be partitioned into activity domains.
- CATCH: Total catch of a stock for a nation or some partition of it, corresponding to SI records. Ideally this can be based on field definitions provided for the CL table. CATCH may be partitioned into activity domains. Catch category is a field on CL, so this supports discards as well, but perhaps discard estimates is better handled in the estimates table.
- ESTIMATES: Total catch composition as total number and mean length or weight of a nations catch or discards of a stock, corresponding to SD records. Total and means may be provided within domains that partition the population of catches into activity and fish domains.

For each of the table above, it should be indicated which fields define the population quantities are provided for. Importantly, all tables should be complete with respect to the population, that is population fields are mandatory (cannot contain missing values), census data should sum to the entire population, and any missing estimates should be indicated by missing values. This is necessary for any combined used of these tables. For all tables, each row should represent a domain in the population, and no domains should overlap with any other. That is every fish in the population of catches should be attributable to exactly one domain.

Extendable file format

In order to meet requirements of extendibility (requirement 4). We can use the approach used for code-lists, and define a vocabulary of available fields for each table. We can specify that these should be stored as csv-files with headers (one file for each of the three tables ESTIMATES, CATCH and EFFORT), and that parsers should assume all values to be NA for vocabulary fields not present in a file. Future versions will be subject to the rule that columns may be added to the format, but not made mandatory, and that existing columns may not be removed in future versions. This allows us to make a large vocabulary, without burdening data providers with columns that are not needed for their particular use, and it allows us to start with a minimal set of fields, and extend upon that when new requirements arise. As mentioned before, this is very similar to how we deal with code-lists. This will also be a key strategy to accommodate requirement 2. If we retain the option of adding fields later on, we can extend what kind of domains can be specified as need arises and we don't have to foresee them all right away.

Columns in the vocabulary may have the following status:

M: mandatory column. Must be present and a valid value must be provided for every domain.

P: optionally present. May be either present or not for a population. If it, is a valid value must be provided for every domain

O: optional. May be present, may have missing values for some domains.

Interoperable format definitions

We will re-use definitions from the following formats which we need to be consistent with for reasons of interoperability:

ICEF: https://www.ices.dk/data/Documents/InterCatch/IC-ExchangeFormat1-0.pdf

CE and CL table: https://github.com/ices-tools-dev/RDBES/tree/master/Documents

BiologicalMeasurements: https://vocab.ices.dk/?ref=1606

Starting vocabulary for ESTIMATES

Value fields

The estimated values reported for a domain may be the total number of fish in that domain or a mean of some observed value. The observed values that are supported by RDBEs are the ones listed in https://vocab.ices.dk/?ref=1606, although means only make sense for numerical variables. Assuming that we retain the option of adding more quantities later one, we will start with defining mean for common observables. In addition, we need some standardized way of reporting uncertainty of the estimate, and some quality indicators (requirement 3). We will attempt with sampling variances, expecting that general formulas for inferring variances of sums and weighted averages of domains are available. We will start off with estimates of standard error, and put in anything we need for compatibility with InterCatch:

FieldType	Name	Description	status
Estimate	totalNumber	Estimated total number of fish in domain.	O
Estimate	totalWeightLive	Estimated total live weight (kg). Live weight defined as WeightLive in https://vocab.ices.dk/?ref=1606	O
Estimate	meanWeightLive	Estimated mean live weight (g). Live weight defined as WeightLive in https://vocab.ices.dk/?ref=1606	O
Estimate	meanTotaLength	Estimated mean total length (cm). Total length defined as TotalLength in https://vocab.ices.dk/?ref=1606	O
Estimate	VarMeanWeightLive	Estimate of sampling variance of the statistic mean live weight (g**2)	O
Estimate	VarMeanTotalLength	Estimate of sampling variance of the statistic mean total length (cm**2)	O
Estimate	VarTotalNumber	Estimate of sampling variance of the statistic total number of fish	O
Estimate	VarTotalWeight	Estimate of sampling variance of the statistic total live weight of fish (kg**2)	O
Sample description	NumSamplesLength	As in intercatch	O
Sample description	NumLngMeas	As in intercatch	O
Sample description	NumSamplesAge	As in intercatch	O
Sample description	NumAgeMeas	As in intercatch	O
Sample description	sampledcatch	As in InterCatch	O

Fish domains

In order to specify fish domains we could define as a part of the vocabulary the list for BVtype-Measured (https://vocab.ices.dk/?ref=1606). These are all observations made on individual specimen and can in principle be used to form fish domains, although it makes most immediate sense for variables that can be treated as discrete variables (Age, Sex, Maturity, Stock, etc.). We will start by including the most sensible. For selected numerical variables (age, length and weight) we will make special provisions so that arbitrary groups can be defined as part of a domain:

Field type	Name	Description	status
Fish domain	MinAge	Lowest age in domain	P
Fish domain	MaxAge	Highest age in domain	P
Fish domain	MinTotalLength	Lowest total length (cm) in domain	P
Fish domain	MaxTotalLength	Highest total length (cm) in domain	P
Fish domain	MinWeightLive	Lowest live weight (g) in domain	P
Fish domain	MaxWeightLive	Highest live weight (g) in domain	P
Fish domain	Sex	As in https://vocab.ices.dk/?ref=1606	P
Fish domain	StockMembership	As Stock in https://vocab.ices.dk/?ref=1606	P

Note that the conventional plusgroup definition can be encoded in this way, setting MinAge=MaxAge for all ages below the plusgroup, and setting MaxAge to a sufficiently high number for the plus group.

Activity domains and population fields

In order to specify activity domains, we could define as part of the vocabulary the columns of CE and CL. We will start with those necessary for compatibility with InterCatch, and the field stockArea, explained below:

Field type	Name	Description	status
Population field	vesselFlagCountry	As in CL and CE	M
Population field	year	As in CL and CE	M
Population field	speciesCode	As in CL	M
Population field	catchCategory	As in CL	M
Activity domain	seasonType	Extendable code list: Month: As in CL and CE Quarter: As in CL and CE	P
Activity domain	seasonValue	Value of season type	O
Activity domain	stockArea	The area defined for the stock. New code list based on ICES_stock.	P
Activity domain	areaType	area: As in CL and CE statisticalRectangle: as in CL and CE	P
Activity domain	areaValue	Value of area type	O
Activity domain	metier6	As in CL and CE	P

In order to represent the coarsest possible spatial resolution for domains, we have included the field stockArea, which is the total area the stock is defined for. A code list can be derived from the ICES stock code list. Together with speciesCode it will also identify the stock domain estimates are provided for.

In addition, we may need some columns for InterCatch interoperability that are not in CL and CE:

Field type	Name	Description	Status
Population field	ReportingCategory	As in InterCatch	M
Activity domain	Fleet	As in InterCatch	P

For full compatibility with InterCatch, the ReportingCategory needs to be included, and it is describing the population that estimates are provided for and it is necessary to know its value to know if it can be combined or compared with other tables, such as CATCH or EFFORT. Because it is describing the population it has been given mandatory status, but we have not given any consideration to whether this classification of the population is something we would like to go forward with, or to what extent it may overlap with other population descriptions, such as CatchCategory. One could consider defining the format without it, and rather if requirement 1 can be reformulated so that InterCatch compatibility is only required for cases when ReportingCategory is 'R' (Reported), which in InterCatch is also used for discards, as long as only reported activity is included in the discards sampling.

The Fleet variable is important as it represents a controlled vocabulary of domain definitions that are controlled by the data users (assessment working groups). We have simply adapted it from InterCatch for compatibility reasons, but have not carefully considered if this way of representing custom domains required by data users is either sufficiently flexible or sufficiently strict. The granularity of these fleets may also be key point of conversation between data providers and data users. In many cases, it is too detailed metier lists that has been forcing low precision estimates and data imputation.

The fleet variable is actually a concatenation of many fleet definitions. This could be made clearer by making one list for each working group, and allow them to control them directly. As long as these are introduced as optionally present fields (P) they can be added to the format as needed, and phase out the need for the field "Fleet" entirely.

Mandatory fields and constraints

Per requirement 2., we would like data requirements to be subject to policy and agreement between stock assessors and regional estimators. We will therefore make mandatory only fields that are necessary in order to clearly describe the population estimated for. As a rule, we make fields optional unless there is a good argument for why it should be mandatory.

All population fields must be mandatory, as the quantity estimated cannot be correctly combined with census data defined without it. This also ensures sum-of-product tests can be performed when the population coincides with what census data is provided (requirement 6). Otherwise, we would like to be as careful as reasonable with mandatory values, leaving those decisions to be agreed between national estimators and stock assessors.

Interpretation of optional fields

For all field types mandatory (M) means that the column must be present, and filled with valid value for every domain. For optional field types there is no difference in interpretation between a column missing, and a column provided with all values missing.

Optionally present fields (P) must either be present or not for a population as defined by the population fields. That is, different countries may use different domains, and a country may use different domains for discards and landings, but cannot provide missing values for P columns when it is included in the domain definition for a population.

For the optional fields, the interpretation of a missing value is different for the different field types. For domain fields (fish domain or activity domain), a missing value means "not elsewhere identified", so that it is not valid to use a missing value unless it is guaranteed that there is no

overlap between that domain and any other domains. For the field types 'Estimate' and 'Sample description', a missing value means NA (not available), which means the value could not be estimated or was not desired to include for that particular domain. Note however that the Sample description fields are usually available even for unsampled domains, while estimates are typically not.

Spatial and temporal fields

When a missing value ("not elsewhere identified") is allowed for more than one field, the same domain may be specified in several different ways. For instance, given that there exist two domains representing February and March. January may be represented either as (Q1, missing month) or the more natural (missing quarter, January) or (Q1, January). In order to compare between the tables, ESTIMATE, EFFORT and CATCH, we need to enforce some consistency in encoding. We did not find a way to both re-use existing spatial and temporal definitions, and meet requirement 5, without introducing NEI values. We therefore decided to rather facilitate checking by introducing some type-value pairs of columns (areaType/areaValue and seasonType/seasonValue). In that way assumptions and agreements about which granularity is used can be easily checked (E.g., check that the same season definition is used within the same metier, etc.). These fields need to be carefully reviewed early on, as it is not clear that it will be easy to add new optional columns without interfering with the validity of earlier versions of the format.

Starting vocabulary for EFFORT and CATCH

The CATCH table should have the same vocabulary for activity domains and population fields as the Estimates table, and the same mandatory status, including those imported from InterCatch. Value fields should include total catch (CATON in InterCatch, CLofficalWeight, ClscientifWeight in CL).

The EFFORT table should contain the same vocabulary for activity domains and population fields as the Estimates table, except any variable that requires knowledge of the catch composition (species, usage, etc.). It should have the same population fields, except species code and catch category. The value fields may include all effort variables in the CE table, and any additional options from the UnitEffort field in ICEF.

A possible exception for the effort table may be metier6. This variable occurs on both CL and CE, and the definition requires knowledge of target species, which is conventionally defined as dominant species in catch for some fisheries but may require a different definition in the EFFORT table as effort that does not result in catch need to be accounted for.

Example

An example for the kind of data formatted by the suggestions in this document is prepared in the file "national_estimates_example_draft.ods".

Tasks identified for elaborating standard TAF data types

- Prepare more formal definition for EFFORT and CATCH
- Evaluate if the data from the LS table in InterCatch needs to be accommodated by RCEF.
- Consider if ReportingCategory should be dealt with differently. Consider overlap with CatchCategory and consider other constraints on conversion from InterCatch.
- Annotate format descriptions with which information needs to be present in order to convert RCEF to ICEF.
- Discuss how the field Fleet is best handled. Referring to questions raised in under the heading "Activity domains and population fields".

- Check if the suggestion for ESTIMATES may be populated without loss of information from InterCatch, and if conversion back to InterCatch can be done when necessary optional fields are filled.
- Check if the EFFORT table may be populated without loss of information from InterCatch, and if conversion back to InterCatch can be done when necessary optional fields are filled.
- Check if the CATCH table may be populated without loss of information from InterCatch, and if conversion back to InterCatch can be done when necessary optional fields are filled.
- Check if the suggestion for ESTIMATES contain sufficient information to infer uncertainty measures for aggregations of the estimates.
- Develop a standard TAF data type for “combined estimates” / stock coordination output.

Annex 3: Structure of the Catch/Effort Overview repository

2023_CatchEffort_RDBES_Combined

```
+-- ARTIFACTS.json
+-- bootstrap
|   +-- DATA.bib
|   \-- initial
|           \-- data
|           +-- 001_data_RDBESmodel
|           |   +-- 001_inputs_data_RDBESmodel
|           |   |   \-- v1.19.4
|           |   |       +-- baseTypes.xsd
|           |   |       +-- RDBES Data Model CL CE.xlsx
|           |   |               +-- RDBES Data Model CS.xlsx
|           |   |               \-- RDBES Data Model VD SL.xlsx
|           |   \-- 002_outputs_data_RDBESmodel
|           |                   \-- v1.19.4
|           |                   +-- RDBES_dataModel.RData
|           |                   \-- RDBES_dataModel.xlsx
|           +-- 002_data_codeLists
|           |   +-- 001_inputs_data_codeLists
|           |   |   \-- v1.19.4
|           |   \-- 002_outputs_data_codeLists
|           |                   \-- v1.19.4
|           |                   +-- RDBES_codeLists.RData
|           |                   \-- RDBES_codeLists.xlsx
|           +-- 003_data_getiso2
|           |   +-- 001_inputs_data_getiso2
|           |   \-- 002_outputs_data_getiso2
|           |                   \-- iso2.csv
|           \-- 004_data_otherCodeLists
|           +-- ASFIS_WoRMS_updt.csv
|           +-- aux_colours.txt
|           +-- aux_colours.txt.bak
|           +-- aux_countries.txt
|           +-- aux_stocks.txt
|           +-- fleetReg
|           |   \-- README.rmd
|           +-- README.rmd
|           +-- Table_Species_Categ.txt
|           +-- UNLOCODE.rData
|           \-- worms_3alpha.csv
+-- data.R
+-- data_files
            |   +-- 001_data_prepareRDBESdatamodel.R
            |   +-- 002_data_prepareRDBEScodelists.R
            |   +-- 003_data_getiso2FAO.R
            |   +-- 004_data_conversion.R
            |   \-- 005_data_preparation.R
+-- LICENSE
+-- model.R
```

```
+-- output.R
+-- report.R
+-- report.Rmd
+-- report_files
|   +-- report_doc.R
|   +-- report_plots.R
|   \-- report_tables.R
+-- utilities.R
\-- utilities_files
            +-- settings
            |   \-- settings.r
            +-- utilities_functions
            |   +-- fun_003_Fleet_registry_analyses.r
            |   +-- fun_barplot_var_by_one_var.r
            |   +-- fun_barplot_var_by_one_var_rmd.R
            |   +-- fun_barplot_var_by_two_var_stacked.r
            |   +-- fun_barplot_var_by_two_var_stacked_rmd.R
            |   +-- fun_choroplethMap_func.R
            |   +-- fun_clean_empty_rows.r
            |   +-- fun_create_CA_from_RDBES.r
            |   +-- fun_create_HH_from_RDBES.r
            |   +-- fun_create_HL_from_RDBES.r
    |   +-- fun_create_RDBES_Id_table.r
            |   +-- fun_create_RDB_CL_from_RDBES.r
            |   +-- fun_create_SL_from_RDBES.r
            |   +-- fun_create_TR_from_RDBES.r
            |   +-- fun_determine_what_to_inset.r
            |   +-- fun_download_data_from_sharepoint.r
            |   +-- fun_FAZ_SEGMENTACAO_COMPRIMENTO_DCF2.R
            |   +-- fun_flowPlot.R
            |   +-- fun_group.R
            |   +-- fun_pointsMap.R
            |   +-- fun_rename.R
            |   +-- fun_riverplotfun.r
            |   +-- fun_scatterpieMap.R
            |   +-- fun_table.R
            |   +-- fun_temporary_adaptation_of_icesVocab_functions.r
            |   \-- fun_theme_flextable.R
            \-- utilities_graphical_parameters
            \-- gp_graphical_parameters.csv
```