Manual for assigning metiers to transversal data

Method developed by RCG ISSG on Metier issues

Date

Contents

Background

In 2018 a DCF Métier workshop was held as a subgroup of the RCG's. Work was done on writing a historical background of the métiers and why they are needed. Participants of the workshop described the national procedures for assigning métiers, and it was clear that different methods and criteria are used for assigning métier codes by the different nations. During the meeting, major issues and possible best practices were discussed, e.g. that the target species assemblage should reflect the fishing intention, and that a métier should be defined for a fishing sequence. Work was started on the reference lists on métiers, species and gears. A general workflow for assigning métiers was developed, and a repository for storing reference lists and scripts was suggested.

A result of the workshop was the consensus regarding the need to standardize and harmonize métier codes and reference tables, as well as the methods to assign métiers to transversal data.

In 2019 the work was followed up by a pan-regional RCG Intersessional Subgroup on Métier Issues that set up a public repository on GitHub for storing reference lists, scripts, métier descriptions and documentation of procedures. The work performed consisted in:

- Suggestion of a system for harmonizing the métier reference lists without overlapping métiers (especially about mesh-size ranges)
- Development of an R-script for assigning métiers to transversal data (tested for the Baltic Sea)
- Development of a template for documenting the method used by countries, including its testing
- Development of a script for making métier descriptions based on data uploaded to the RDB
- Agree on and test a reference list of individual species to be included into species groups
- Test and evaluate the impact of using value vs. weight of landings as metric for assigning target species assemblage groups.

The RCG's in 2019 recommended that the work was to be followed up and continued in 2020, with a stronger focus on an operational métier list and R-script to assign métiers. This includes testing reference tables and R-script and making sure that all relevant métiers and reference codes are included.

The reports from the 2018 workshop and the 2019 and 2020 RCG intersessional work can be found in the RCG github repository.

This document is a practical manual for using the scripts and reference codes created by the ISSG on Métier issues. First, the input data format for transversal data is described in Section 2. The reference lists used in the R-script for métier codes, species codes and area codes are described in Section 3. In Section 4 the R-script is described and it is explained how to use it.

Input format for transversal data

The script developed for assigning métier codes to transversal data reads a ".csv" file with the input data format described below.

Column Name	Description	Format	Example
Country	Country code	ISO 3166-1 alpha-3 codes	POL
year	Year	AAAA	2019
vessel_id	Vessel id		AAA-1
vessel_length	Vessel length	meters	12.01
trip_id	Trip identifier		POLAAA1201806100325
haul_id	Haul identifier. Optional.		POLAAA1201806100325_1
fishing_day	Fishing date	DD-MM-AAAA	10-06-18
area	FAO area code. To be used with the area reference list to identify the RCG.		27.3.d.25
ices_rectangle	ICES rectangle code. Optional outside ICES area.		37G5
gear	Gear code	FAO ISSCFG code	OTB
mesh	Mesh size	integer	100
selection	Selection panel code. A combination of selection panel code number and selection panel mesh size.	0: No selection device, 1: Selection panel, 2: Grid, 3: T90	1_120
registered_target_assemblage	If the target assemblage code is registered in the logbooks it can be entered here. The field is optional		DEF
FAO_species	Species code	FAO ASFIS Code	COD
Metier_level_6	Should not be filled in. The resulting mÃ©tier will be inserted into this column.		
measure	If the target assemblage calculation should be based on weight or value.		weight or value
KG	Kg weight of landing		Precision: 2 digits after the decimal
EUR	Value of landing in euro		Precision: 2 digits after the decimal

An example of the input data format can be found here: https://github.com/ices-eg/RCGs/blob/master/Metiers/Metier_data_format_Example_test_input.csv

Reference lists

Métier list

The R-script downloads the métier list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/RDB_ISSG_Metier_list.csv. It is also available in excel format. The list has information on RCG (BALT/LDF/MBS/NAtl/NSEA) and the métier code (Metier_level6). In some cases, there is a start and end year for the métier code due to legislation, and they are marked in the "start_year" and "end_year" columns. The column "old_code" is a reference to the codes used in the previous métier list from the RDB. If they are marked with red in the excel file, the code has changed in the new métier list. The column "Use_by_country_in_RDB" lists the countries that have uploaded data with the métiers in the RDB for the 2009-2017 periods, and in the column "Total_n_trips_RDB_2009t2017", presents the sum of the total number of trips with the old codes uploaded to the RDB for 2009 to 2019.

Species list

The R-script uses the species list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/Metier%20Subgroup%20Species%202020.xlsx The list is based on the FAO ASFIS species list with the FAO species codes in column A and associated information in columns B to G. The columns H to T reflect the work on grouping the species codes in the métier group, the R data package, the RCM 2007 and what has been done within countries. Columns U, V and W contains the latest proposal on species groups at three levels and is used by the R-script. Grouping 1 in column V groups the species into Crustaceans (CRU), Molluscs (MOL), Finfish (FIF), Seaweeds (SWD) and Miscellaneous (MIS) Grouping 2 in column W splits the finfish category up into Anadromous (ANA), Catadromous (CAT), Demersal (DEF), Small pelagic fish (SPF), Large pelagic fish (LPF) and Freshwater species (FWS). Also Cephalopods (CEP) are separated from other molluscs. Grouping 3 - DWS identifies deep water species from the regulation (EC) 2016/2336. The Groupings 2 and 3 are used by the script, but if a country needs to group species differently, it can be done at the national level by including the changes in the R-script.

Area list

The script downloads the area list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/AreaRegionLookup.csv. This list includes areas on the levels Area, SubArea, Division and SubDivision. It links the area codes to the RCG code that is used in the métier list.

Gear list

The R-script downloads the area list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/ Code-ERSGearType-v1.1.xlsx.

Selection panels

The last part of the metier codes define the selection panels. They are often _0_0 in the last part of the metier code. Selection panel codes have been defined in some of the métiers and can be written in the input data format in the "selection" column.

The first number define the type of selection panel:

Code number Selection panel type

0 No selection device 1 Selction panel 2 Grid 3 T90 4 There is both selection device and escape window. Specify the smallest mesh size.

Métier assignment procedure

An R-script was developed and tested in 2019 for assigning métiers in the Baltic Sea, which has been further developed in 2020 to cover more areas and details. The R-script is available at https://github.com/ices-eg/RCGs/blob/master/Metiers/Scripts/script_metiers_test.R and functions are available at https://github.com/ices-eg/RCGs/tree/master/Metiers/Scripts/Functions.

Assigning "metier_level_6"

The script loads the transversal data in the format described in Section 2, the reference lists described in Section 3, and assigns the RCG name to the input data and the species target assemblage group. It is possible to define the sequence, which is the level that the métier is assigned to. The default is by following columns: Country, year, vessel id, vessel length, trip id, haul id, fishing day, area, ices rectangle, gear, mesh size, selection panel an registered target species assemblage The sequence can be changed to the preferred level. The species target assemblage group is determined for each sequence. If there is more than 8% deep water species, it is assigned to the DWS target species assemblage group. The level 6 métier code is assigned to the "column metier_level_6" by combining the information of the RCG, year, gear, target assemblage, mesh size, selection panel type and selection panel mesh size, with the métier list.

National corrections

It is the idea that the script can be used as the core method to assign the métiers, but that countries could add national corrections. These can include, for example, corrections of a species grouping if the species is fished within another métier nationally or corrections of gear codes (e.g. grouping or recoding imprecise gear codes (GN, TB)). It is encouraged that the countries upload their script with national corrections to GitHub under https://github.com/ices-eg/RCGs/tree/master/Metiers/Documentation_by_MS along with the documentation of the methods and for other countries to use same methods in their scripts if relevant, e.g. for defining selection panels etc. The lines with national corrections can be marked with #Country code. The scripts uploaded don't need to be final, they can be updated as they improve.

Script looking at vessel patterns for assigning métiers

A function in the métier allocation script has been developed that takes the vessel pattern into account. This function will output a level 5 métier code in the column "metier_level_5". This script identifies the main métiers used by a vessel using a stepwise procedure that is described in Section 5.

Results

The resulting file, which is the input data with results added is saved as a ".csv" file. It is possible to specify the path. The results are also summed up and saved in an excel file that gives an overview of the métiers allocated.

R script

The script developed to assign metiers based on several variables is called script_metiers_test.R.

Prerequisites

The packages required to run the script are (they can be downloaded from the CRAN, https://cran.r-project.org/web/packages/available_packages_by_name.html),

- stringr
- data.table
- openxlsx
- purr

as well as auxiliary information described in sections XXX and a set of functions developed to facilitate the readability of the script. These functions are descripted below in detail.

- loadInputData.R : reads the Input file provided it's in csv format
- loadAreaList.R : reads the RCG area file from the Github repository
- loadSpeciesList.R : reads the RCG species file from the Github repository
- loadMetierList.R : reads the RCG metier file from the Github repository
- loadGearList.R : reads the RCG gear file
- getMeasure.R : determines the if the sequence factor is weight or value.
- getMetier.R : assigns the metier level 6
- getMetierLvl5FromPattern.R : assigns metier level 5

Data

The data used as an input should be a csv file format as described in detail in section 2 of the report (Input format for transversal data). The example data set is shown below for clarity.

Country	year	vessel_id	vessel_length	trip_id	haul_id	fishing_day	area	ices_rectangle	gear	mesh	selection	registered_target_assemblage	FAO_species	metier_level_6	measure	KG	EUR
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	HXC	NA	weight	400.00	2500.0000
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	SHO	NA	weight	200.00	400.0000
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	HER	NA	weight	500.00	475.0000
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	SPR	NA	weight	1800.00	1710.0000
POL	2018	AAA-1	15.00	POLAAA1201807011230	NA	02-07-2018	27.3.d.25	37G5	OTB	110	NA	NA	COD	NA	weight	500.00	NA
POL	2018	AAA-1	15.00	POLAAA1201807011230	NA	02-07-2018	27.3.d.25	37G5	OTB	110	NA	NA	FLE	NA	weight	300.00	NA
POL	2018	AAA-1	15.00	POLAAA1201807011230	NA	03-07-2018	27.3.d.25	37G5	OTB	110	NA	NA	PLE	NA	weight	100.00	NA
POL	2018	ZZZ-2	12.01	POLZZZ2201801250940	NA	25-01-2018	27.3.d.25	38G5	OTB	105	1_120	NA	COD	NA	value	146.00	138.7000
POL	2018	ZZZ-2	12.01	POLZZZ2201801250940	NA	25-01-2018	27.3.d.25	38G5	OTB	105	1_120	NA	FLE	NA	weight	1500.00	1425.0000
POL	2018	ZZZ-2	12.01	POLZZZ2201801250940	NA	25-01-2018	27.3.d.25	38G5	OTB	105	1_120	NA	PLE	NA	weight	50.00	47.5000
POL	2018	BBB-3	11.99	POLBB3123456	NA	21-06-2018	27.3.d.24	37G4	GNS	220	NA	NA	COD	NA	value	146.25	138.9375
POL	2018	BBB-3	11.99	POLBB3123456	NA	21-06-2018	27.3.d.24	37G4	GNS	220	NA	NA	FLE	NA	weight	25.00	23.7500
POL	2018	BBB-3	11.99	POLBB3123456	NA	21-06-2018	27.3.d.24	37G4	GNS	220	NA	NA	PLE	NA	weight	25.00	23.7500
POL	2018	BBB-3	11.99	POLBB3123456	NA	21-06-2018	27.3.d.24	37G4	GNS	220	NA	NA	TUR	NA	weight	200.00	190.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	14-04-2020	27.3.d.26	39G9	GNS	110	NA	FWS	FLE	NA	weight	50.00	100.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	14-04-2020	27.3.d.26	39G9	GNS	110	NA	FWS	COD	NA	weight	100.00	200.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	14-04-2020	27.3.d.26	39G9	GNS	110	NA	FWS	FPE	NA	weight	50.00	100.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	15-04-2020	27.3.d.26	39G9	GNS	110	NA	DEF	FLE	NA	weight	200.00	400.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	15-04-2020	27.3.d.26	39G9	GNS	110	NA	DEF	COD	NA	weight	80.00	160.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	15-04-2020	27.3.d.26	39G9	GN	1	NA	DEF	FPE	NA	weight	300.00	500.0000
POL	2010	ZZZ-2	12.01	POLZZZ2201001250940	NA	17-03-2010	27.3.d.25	38G5	OTB	120	NA	NA	COD	NA	value	50.00	NA
POL	2010	ZZZ-2	12.01	POLZZZ2201001250940	NA	17-03-2010	27.3.d.25	38G5	OTB	120	NA	NA	FLE	NA	weight	60.00	NA
POL	2010	ZZZ-2	12.01	POLZZZ2201001250940	NA	17-03-2010	27.3.d.25	38G5	OTB	120	NA	NA	PLE	NA	weight	70.00	NA
POL	2017	GDY-555	40.00	POLGDY555201701011230	NA	02-01-2017	87.3.3	NA	OTM	50	NA	NA	CJM	NA	weight	10000.00	NA
POL	2017	GDY-555	40.00	POLGDY555201701011230	NA	02-01-2017	87.3.3	NA	OTM	50	NA	NA	MAS	NA	weight	8000.00	NA
POL	2020	DDD-4	28.50	POLDD9898	NA	28-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	FLE	NA	weight	40.00	120.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	29-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	FLE	NA	weight	80.00	240.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	30-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	FLE	NA	weight	60.00	180.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	01-05-2020	27.3.d.26	39G9	OTB	115	NA	NA	FLE	NA	weight	70.00	210.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	02-05-2020	27.3.d.26	39G9	OTB	115	NA	NA	FLE	NA	weight	100.00	300.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	02-05-2020	27.3.d.26	39G9	OTB	115	NA	NA	COD	NA	weight	20.00	100.0000
POL	2020	DDD-4	28.50	POLDD9898	NA	08-05-2020	27.3.d.26	39G9	XYZ	115	NA	NA	YYY	NA	weight	20.00	100.0000

Code

The first step of the R-script splits the "selection" column information, if any, between the selectivity device code *(e.g. 1)* and the selectivity device mesh size *(e.g. 120)*.

```
input.data[, `:=`(c("selection_type", "selection_mesh"), data.table(str_split_fixed(selection,
    "_", 2)))]

input.data[, `:=`(selection_type, ifelse(selection_type == "",
    NA, selection_type))]

input.data[, `:=`(selection_mesh, ifelse(selection_mesh == "",
    NA, selection_mesh))]
```

selection	selection_type	selection_mesh
1_120	1	120
1_120	1	120
1_120	1	120

Then the R-script assigns a RCG area code based on the "area" column information provided in the input data and the area reference list.

```
input.data <- merge(input.data, area.list, all.x = T, by = "area")
input.data[is.na(RCG) & substr(area, 1, 2) %in% c("31", "34",
    "41", "47", "51", "57", "58", "87"), `:=`(RCG, "LDF")]
input.data[is.na(RCG) & substr(area, 1, 2) == "37", `:=`(RCG,
    "MED")]
```

area	RCG	Description
27.2.a	NSEA	North Sea and Eastern Arctic
27.3.d.24	BALT	Baltic Sea
27.3.d.25	BALT	Baltic Sea
27.3.d.26	BALT	Baltic Sea
87.3.3	LDF	Distant Waters

The next step of the R-script is to assign the target assemblage group of species associated with the species landed with information provided in the "FAO_species' column and the species reference list. The possible belonging of the species to the Deep Water Species (DWS) group is also calculated.

```
input.data <- merge(input.data, species.list, all.x = T, by = "FAO_species")
```

FAO_species	species_group	dws_group
CJM	SPF	NA
COD	DEF	NA
FLE	DEF	NA
FPE	FWS	NA
HER	SPF	NA
HXC	LPF	DWS
MAS	SPF	NA
PLE	DEF	NA
SHO	DEF	DWS
SPR	SPF	NA
TUR	DEF	NA
YYY	NA	NA

A set of variables *(combination)* are necessary to define the fishing sequence level at which the métier will be calculated and assigned. Total tonnage landings in weight and in value is also calculated by target assemblage group of species and fishing sequence. The default fishing sequence is shown below, however it can be changed to reflect the national combination of variables, that is up to the country to consider.

```
sequence.def <- c("Country", "year", "vessel_id", "vessel_length",
    "trip_id", "haul_id", "fishing_day", "area", "ices_rectangle",
    "gear", "mesh", "selection", "registered_target_assemblage")

# Calculate group totals for each sequence
input.data[, `:=`(seq_group_KG = sum(KG, na.rm = T), seq_group_EUR = sum(EUR,
    na.rm = T)), by = c(sequence.def, "species_group")]
```

Below you can see the calculated totals in the example dataset with the default sequence.

Country	year	vessel_id	vessel_length	trip_id	haul_id	fishing_day	area	ices_rectangle	gear	mesh	selection	registered_target_assemblage	species_group	**seq_group_KG**	**seq_group_EUR**
POL	2017	GDY-555	40.00	POLGDY555201701011230	NA	02-01-2017	87.3.3	NA	OTM	50	NA	NA	SPF	**18000.00**	**0.0000**
POL	2018	BBB-3	11.99	POLBB3123456	NA	21-06-2018	27.3.d.24	37G4	GNS	220	NA	NA	DEF	**396.25**	**376.4375**
POL	2018	AAA-1	15.00	POLAAA1201807011230	NA	02-07-2018	27.3.d.25	37G5	OTB	110	NA	NA	DEF	**800.00**	**0.0000**
POL	2018	ZZZ-2	12.01	POLZZZ2201801250940	NA	25-01-2018	27.3.d.25	38G5	OTB	105	1_120	NA	DEF	**1696.00**	**1611.2000**
POL	2010	ZZZ-2	12.01	POLZZZ2201001250940	NA	17-03-2010	27.3.d.25	38G5	OTB	120	NA	NA	DEF	**180.00**	**0.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	14-04-2020	27.3.d.26	39G9	GNS	110	NA	FWS	DEF	**150.00**	**300.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	15-04-2020	27.3.d.26	39G9	GNS	110	NA	DEF	DEF	**280.00**	**560.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	02-05-2020	27.3.d.26	39G9	OTB	115	NA	NA	DEF	**120.00**	**400.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	28-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	DEF	**40.00**	**120.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	29-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	DEF	**80.00**	**240.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	30-04-2020	27.3.d.26	39G9	OTB	115	NA	NA	DEF	**60.00**	**180.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	01-05-2020	27.3.d.26	39G9	OTB	115	NA	NA	DEF	**70.00**	**210.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	14-04-2020	27.3.d.26	39G9	GNS	110	NA	FWS	FWS	**50.00**	**100.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	15-04-2020	27.3.d.26	39G9	GN	1	NA	DEF	FWS	**300.00**	**500.0000**
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	SPF	**2300.00**	**2185.0000**
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	LPF	**400.00**	**2500.0000**
POL	2018	AAA-1	15.00	POLAAA1201807011230	NA	03-07-2018	27.3.d.25	37G5	OTB	110	NA	NA	DEF	**100.00**	**0.0000**
POL	2018	AAA-1	15.00	POLAAA1201806100325	NA	10-06-2018	27.2.a	NA	OTB	100	NA	NA	DEF	**200.00**	**400.0000**
POL	2020	DDD-4	28.50	POLDD9898	NA	08-05-2020	27.3.d.26	39G9	XYZ	115	NA	NA	NA	**20.00**	**100.0000**

The function "getMeasure.R" is then used to determine by fishing sequence the dominant factor *(weight or value)*, as stated in the measure column, to be considered to define the principal target assemblage group of species for the fishing sequence considered. If at least one species in the fishing sequence presents the modality "value", then the reference measure becomes the value for the whole fishing sequence.

```
input.data[, `:=`(seq_measure = getMeasure(measure)), by = sequence.def]
```

FAO_species	area	**measure**	RCG	species_group	**seq_group_KG**	**seq_group_EUR**
CJM	87.3.3	**weight**	LDF	SPF	**18000.00**	**0.0000**
COD	27.3.d.24	**value**	BALT	DEF	**396.25**	**376.4375**
COD	27.3.d.25	**weight**	BALT	DEF	**800.00**	**0.0000**
COD	27.3.d.25	**value**	BALT	DEF	**1696.00**	**1611.2000**
COD	27.3.d.25	**value**	BALT	DEF	**180.00**	**0.0000**
COD	27.3.d.26	**weight**	BALT	DEF	**150.00**	**300.0000**

The next step of the R-script assigned by fishing sequence the principal/main target assemblage group of species based on the highest tonnage total landings in weight or in value observed by fishing sequence. It means that the "species group" column information could then differ from "seq_dom_group" for the fishing sequences where more than one "species_group" has been landed when only one unique "seq_dom_group" will be assigned to the whole fishing sequence.

```
input.data[seq_measure == "weight", `:=`(seq_dom_group = species_group[which.max(seq_group_KG)]),
    by = sequence.def]
input.data[seq_measure == "value", `:=`(seq_dom_group = species_group[which.max(seq_group_EUR)]),
    by = sequence.def]
```

The DWS rule is also applied when more than 8% of the species landed in weight, in the fishing sequence considered, belong to the deep water species group. In this case, the DWS target species assemblage group is assigned to the whole fishing sequence as the dominant/principal target assemblage group of species in the *"seq_dom_group"* column.

```
input.data[dws_group == "DWS", `:=`(seq_DWS_kg, sum(KG, na.rm = T)),
    by = c(sequence.def, "dws_group")]
input.data[, `:=`(seq_total_kg, sum(KG, na.rm = T)), by = sequence.def]
input.data[, `:=`(seq_DWS_perc, ifelse(is.na(seq_DWS_kg), 0,
```

```
    seq_DWS_kg/seq_total_kg) * 100)]
input.data[, `:=`(seq_DWS_perc, max(seq_DWS_perc)), by = sequence.def]
input.data[seq_DWS_perc > 8, `:=`(seq_dom_group, "DWS")]
```

FAO_species	area	measure	RCG	species_group	seq_group_KG	seq_DWS_perc	seq_dom_group
HER	27.2.a	weight	NSEA	SPF	2300	20.68966	DWS
HXC	27.2.a	weight	NSEA	LPF	400	20.68966	DWS
SHO	27.2.a	weight	NSEA	DEF	200	20.68966	DWS
SPR	27.2.a	weight	NSEA	SPF	2300	20.68966	DWS

The function "getMetier.R" first checks if the user has provided in the input data and for the fishing sequence considered a target assemblage group of species. If it occurs it gives it priority over the dominant assemblage group of species before calculated.

Then, the R-script conditionally assigns the **métier level 6** following a stepwise procedure:

1. The the métier is firstly assigned to the fishing sequence considered based on the combination of the RCG code, the year, the gear, the dominant target assemblage group of species, the mesh size, the selectivity device and the mesh size of the selectivity device faced to the métier reference list. The métiers with 0 or >0 mesh size are disregarded from the métier reference list.
2. If the métier is not assigned, the next step is to disregard the selectivity device code and mesh size of the selectivity device from the combination and face again the combination of the RCG code, the year, the gear, the dominant target assemblage group of species and the mesh size to métier reference list in order to possibly assign the corresponding métier to the fishing sequence considered. The métiers with 0 or >0 mesh size are also disregarded from the métier reference list.
3. If the métier is not assigned, the next step is to disregard the mesh size from the combination and face again the combination of the RCG code, the year, the gear and the dominant target assemblage group of species to the metier reference list limited to the codes with 0 and >0 mesh size codification in order to possibly assign the corresponding métier to the fishing sequence considered.
4. If the métier at this step is still not assigned *(i.e. NA)*, then the MIS_MIS_0_0_0 métier is assigned to the fishing sequence considered.

```
input.data$metier_level_6 <- NA
input.data[, `:=`(metier_level_6, as.character(pmap(list(RCG,
    year, gear, registered_target_assemblage, seq_dom_group,
    mesh, selection_type, selection_mesh), function(r, y, g,
    t, d, m, st, sm) getMetier(r, y, g, t, d, m, st, sm))))]
```

RCG	year	gear	registered_target_assemblage	seq_dom_group	mesh	selection_type	selection_mesh	metier_level_6
LDF	2017	OTM	NA	SPF	50	NA	NA	OTM_SPF_32-69_0_0
BALT	2018	GNS	NA	DEF	220	NA	NA	GNS_DEF_>=157_0_0
BALT	2018	OTB	NA	DEF	110	NA	NA	OTB_DEF_105-115_1_110
BALT	2018	OTB	NA	DEF	105	1	120	OTB_DEF_105-115_1_120
BALT	2010	OTB	NA	DEF	120	NA	NA	OTB_DEF_>=120_3_120
BALT	2020	GNS	FWS	DEF	110	NA	NA	GNS_FWS_>0_0_0

The metier vessel pattern is used to assign **metier level 5**. The code below identifies the main metiers used by a vessel by taking into account the combination of gear and target assemblage. The column seq_no_lvl5 shows the number of times the metier level 5 was used by the respective vessel in the same year. The percentage (seq_perc_lvl5 column) is used to assign the dominant metier level 5 for each vessel.

```
input.data[, `:=`(metier_level_5, paste(gear, ifelse(is.na(registered_target_assemblage),
    seq_dom_group, registered_target_assemblage), sep = "_"))]
pattern <- unique(input.data[, .SD, .SDcols = c(sequence.def,
    "metier_level_5")])
pattern <- pattern[, .(seq_no_lvl5 = .N), by = .(year, vessel_id,
    metier_level_5)]
pattern[, `:=`(seq_perc_lvl5, seq_no_lvl5/sum(seq_no_lvl5, na.rm = T) *
    100), by = .(year, vessel_id)]
pattern <- pattern[!is.na(metier_level_5)]
input.data <- merge(input.data, pattern, all.x = T, by = c("year",
    "vessel_id", "metier_level_5"))
```

year	vessel_id	metier_level_5	seq_no_lvl5	seq_perc_lvl5
2017	GDY-555	OTM_SPF	1	100.00000
2018	BBB-3	GNS_DEF	1	100.00000
2018	AAA-1	OTB_DEF	2	66.66667
2018	ZZZ-2	OTB_DEF	1	100.00000
2010	ZZZ-2	OTB_DEF	1	100.00000
2020	DDD-4	GNS_FWS	1	11.11111
2020	DDD-4	GNS_DEF	1	11.11111
2020	DDD-4	OTB_DEF	5	55.55556
2020	DDD-4	GN_DEF	1	11.11111
2018	AAA-1	OTB_DWS	1	33.33333
2020	DDD-4	XYZ_NA	1	11.11111

If a "rare" metier is assigned based on one different catch composition or a mis-assigned métier code, it will be set to one of the main métiers of the vessel belonging to this pattern. The percentage threshold for defining rare metiers can be set in the script. The default is 13, meaning that if there is less than 13 percent of the sequences that has a level 5 allocated, it is considered rare and will get the suffix "rare_" in the column "metier_level_5". If possible, the script will suggest another major metier used by the vessel to be used instead. If it is different from the méetier code assigned it will have the suffix "pattern_".

```
rare.threshold <- 13
input.data[seq_perc_lvl5 < rare.threshold, `:=`(metier_level_5,
    NA)]
pattern <- pattern[seq_perc_lvl5 >= rare.threshold]
pattern[, `:=`(c("gear", "target_assemblage"), data.table(str_split_fixed(metier_level_5,
    "_", 2)))]
pattern <- merge(pattern, gear.list, all.x = T, by.x = "gear",
    by.y = "gear_code")
input.data <- merge(input.data, gear.list, all.x = T, by.x = "gear",
    by.y = "gear_code")
input.data[is.na(metier_level_5), `:=`(metier_level_5, as.character(pmap(list(vessel_id,
    year, gear, gear_group, registered_target_assemblage, seq_dom_group),
    function(v, y, g, gg, rt, d) getMetierLvl5FromPattern(v,
        y, g, gg, rt, d))))]
```

vessel_id	year	gear	registered_target_assemblage	seq_dom_group	metier_level_5
DDD-4	2020	GN	DEF	FWS	pattern_OTB_DEF
BBB-3	2018	GNS	NA	DEF	GNS_DEF
BBB-3	2018	GNS	NA	DEF	GNS_DEF
BBB-3	2018	GNS	NA	DEF	GNS_DEF
BBB-3	2018	GNS	NA	DEF	GNS_DEF
DDD-4	2020	GNS	DEF	DEF	pattern_OTB_DEF

Workflow diagramm