Manual for assigning metiers to transversal data

Method developed by RCG ISSG on Metier issues

Date

Contents

1   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.

2 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	Example
Country (M)	Country code (ISO 3166-1 alpha-3 codes)	POL
year (M)	Year (format AAAA)	2019
vessel id (M)	Vessel id	AAA-1
vessel length (M)	Vessel length (in meters)	12.01
trip id (M)	Trip identifier	POLAAA1201806100325
haul id (O)	Haul identifier.	POLAAA1201806100325_1
fishing day (M)	Fishing date (format DD-MM-AAAA)	43261
area (M)	FAO area code. This information will be used to identify the RCG based on the area reference list.	27.3.d.25
ices rectangle (O)	ICES rectangle code.	37G5
gear (M)	Gear code (FAO ISSCFG code)	OTB
gear FR (O)	Gear code (FAO ISSCFG code) from fleet register. The first/main gear, used for metier estimation if the gear code is not known from logbooks	OTB
mesh (M)	Mesh size (integer)	100
selection (O)	Selection panel code which combines the selection panel code number (see allowed reference code hereunder) and the selection panel mesh size in the second part of the codification. Allowed selection panel code numbers: 0: No selection device 1: Selection panel 2: Grid 3: T90 4: For the situation when there is both a selection device and escape window (in this case the smallest mesh size should be specified in the second part of the codification). e.g. possible options for Kattegat: 1_120, 1_140, 1_180, 1_270 or 1_300.	1_120
registered target assemblage (O)	If the target assemblage of species code is registered in the logbooks it can be entered here and the metier level6 will be calculated taking it into account.	DEF
FAO species (M)	Species code (FAO ASFIS Code)	NEP
Metier level 6	Should not be filled in (put it as NA or empty field). The resulting métier calculated by the algorithm will be inserted into this column.	

Column Name	Description	Example
measure (M)	In order to decide how the principal target assemblage of species should be calculated on the basis 1) of the weight or 2) of the value. As soon as, for the fishing sequence considered, there is one entry with the modality value provided, then all the calculation algorithm will be based on value for it.	Accepted values: weight or value
KG (M)	Total tonnage of landing in weight (in Kg)	Real number with a precision: of 2 digits after the decimal
EUR (O)	Total tonnage of landing in value (in euro)	Real number with a precision: of 2 digits after the decimal

Column Name: Mandatory/Optional Field

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

3 Reference lists

3.1 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 métiers reference list at DCF level6 (see "Metier_level6" column), to be considered by the algorithm, has been established by RCG (see "RCG" column): Baltic (BALT), Long Distance Fisheries (LDF), Mediterranean and Black Sea (MBS), North Atlantic (NAtl) and North Sea and Eastern Arctic (NSEA). In some cases, the reference métier codes are dated especially when the reference métier code is allowed only during a fixed time period (see columns "Start_year" and "End-year"). This could be due to legislation.

The column "old_code" is a reference to the correspondent 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 considered métiers in the RDB for the 2009-2017 periods and the column "Total_n_trips_RDB_2009t2017" present the sum of the total number of trips with the old codes uploaded to the RDB for 2009 to 2017.

3.2 Species list

The R-script dowloads the species list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/Metier%20Subgroup%20Species%202020.xlsx

The species list is a subset from the ASFIS list of species maintained by FAO (http://www.fao.org/fishery/collection/asfis/en) and include, as far as possible, all the commercial species landed in the RCG fishing areas. FAO species code for the commercial species considered is specified in column A and for each of them some complementary information are available in columns B to G (e.g. "Scientific name", Taxocode", "ISSCAAP description", . . .). The columns H to T reflect the different ways on grouping the species codes in target assemblage of species groups developed for different purposes: R data package, fishPi EU project, RCM 2007, DWS regulation and what has been done/referenced within countries. In the end, columns

3

U, V and W contain the latest proposal on species groups at three different levels of aggregation. **These aggregations are used by the R-script:**

Grouping 1 in column U aggregate the commercial species into the following target assemblage of species: Crustaceans (CRU), Molluscs (MOL), Finfishes (FIF), Seaweeds (SWD) and Miscellaneous (MIS).

Grouping 2 in column V disaggregate the finfishes category into the following target assemblage of species: Anadromous species (ANA), Catadromous species (CAT), Demersal fish (DEF), Small pelagic fish (SPF), Large pelagic fish (LPF) and Freshwater species (FWS). It disaggregates also the molluscs category into the following target assemblage of species: Cephalopods (CEP) and Other molluscs (MOL).

Grouping 3 - DWS identifies deep water species from the regulation (EC) 2016/2336.

Groupings 2 and 3 constitute the reference basis used by the script and the algorithm to aggregate commercial species into target assemblage of species. Nevertheless and if needed, countries are free to group species differently at national level by including the changes in the R-script.

3.3 Area list

The R-script downloads the area list from https://github.com/ices-eg/RCGs/blob/master/Metiers/Reference_lists/AreaRegionLookup.csv.

Area, SubArea, Division, SubDivision and Unit FAO level (http://www.fao.org/fishery/area/search/en) fishing area codes are listed in the reference list with their links to the RCG code (same code used in the métier reference list). All the FAO fishing areas possible codification at the different levels are listed except the FAO fishing areas codes at "SubArea level" (e.g. 27.5) or at "Area level" (e.g. 27) which encompass several RCGs. Therefore, for FAO area 27 it is promoted to provide, as a minimum level, FAO area at "Division level" in the input data.

3.4 Gear list

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

ISSCFG FAO fishing gear classification codes (http://www.fao.org/cwp-on-fishery-statistics/handbook/capture-fisheries-statistics/fishing-gear-classification/en/) are detailed in column A and for each of them: 1) a description of the code is provided in column B and 2) in column C an aggregation by group of fishing gears (e.g. seine nets) is proposed; such aggregation is possibly used in some specific steps of the algorithm.

3.5 Selection panels

The last part of the metier codes define the selection panels. When no selection devices are used then "_0_0" codification should be used in the last part of the metier code. For métiers where selection devices are used, following selection panel codes number have been referenced. They have to be filled in the input data format in the "selection" column with, in the second part of the codification, the precision of their mesh size. For the situation when there is both a selection device and escape window (selection panel code number = 4) then the smallest mesh size should be specified in the second part of the codification.

Following selection panel code number have been referenced for the different type of selection panel:

Code number	Selection panel type
0	No selection device
1	Selection panel
2	Grid
3	T90
4	For the situation when there is both a selection device and escape window.

4 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 input transversal data in the format described in Section 2, the reference lists described in Section 3, and assigns a RCG code based on the fishing area detailed in the input data *(see "area" column)* and a target assemblage group of species based on the specie landed detailed in the input data *(see "FAO_species" column)* to each row of the input data. The default fishing sequence, constituting the level at which the metier level6 will be calculated and assigned, is defined by the following columns: *"Country, year, vessel id, vessel length, trip id, haul id, fishing day, area, ices rectangle, gear, mesh size, selection panel & registered target species assemblage".*

Nevertheless, countries are free to define the level of fishing sequence they want to consider nationally. For example, countries could take the decision to not consider fishing "ices rectangle area" when defining the level of fishing sequence at which they want to calculate and assign metier level 6. Therefore they could consider only the following columns: *"Country, year, vessel id, vessel length, trip id, haul id, fishing day, gear, mesh size, selection panel & registered target species assemblage".*

The principal target assemblage group of species is then calculated by fishing sequence at the level defined before. The principal target assemblage group of species is determined based on the maximum *(regarding all the target assemblage group of species landed for the fishing sequence considered)* total tonnage of landing calculated by fishing sequence in weight *(if all the "measure" column for the fishing sequence considered is filled out with "weight" modality)* otherwise in value. Specific algorithm allocates the DWS target species assemblage group as principal target assemblage group of species if deep water species represent more than 8% *(according to the Deep Water Regulation (EU) 2016/2336)* of the total tonnage landed in weight.

Then the combination of the RCG code, the year, the gear, the principal target assemblage group of species calculated before and the mesh size is merged to the métier reference list. The correspondent codification, if available, is therefore assigned to the fishing sequence considered and is detailed in the column *"Metier_level_6"*. In the case there is no correspondence with the métier reference list, then no métier level6 is assigned at this step to the fishing sequence and the column *"Metier_level_6"* is left blank or empty.

National corrections

The R-script has been developed in order to improve normalization between countries to assign métier level6 to their national transversal data. It meets a need identified during the 2018 DCF métier workshop which emphasizes the necessity to standardize and harmonize the methods to assign métiers to transversal data between countries. It seeks to be used as the core method shared between countries to assign métiers to fishing sequences. Nevertheless , countries could have some national specificities which have to be taken into account when assigning métiers. These national corrections could be added into the R-script *(for example:*

corrections of a species grouping if the species are fished within another métier nationally or corrections of gear codes (e.g. grouping or recoding imprecise gear codes (GN, TB)).

Countries are encouraged to 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. Other countries could then use similar methods in their scripts if relevant, e.g. for defining selection panels etc. The lines with national corrections should 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 R-script enables to refine the métiers firstly assigned by the general algorithm to take into account the "vessel patterns" i.e. dominant/main metiers used by vessels. This new step contributes to 1) limit the multiplication of métiers calculated *(focus on the yearxvessel' main metiers avoiding "rare" métiers firstly assigned by the general algorithm), 2)* enables the metier calculation when gear is missing, inaccurate/imprecise (e.g. TB, FIX, GN) or wrong *(e.g. typing error, reporting error, . . .) and 3)* enables the metier calculation for fishing sequences without any catches *("zero catches" fishing sequences).* This function will output a revised level 5 métier code in the column "metier_level_5" and this information could be used to assign a reviewed level 6 métier to the fishing sequence *(up to the end-user).*

The stepwise procedure developed is described in detail in Section 5.

Results

The resulting file, which is the input data with algorithm' outputs *(i.e. outputs resulting from the implementation of the algorithm)* added *(e.g. "Metier_level6")* is saved as a ".csv" file into a path which has to be specified. The results are also summed up and saved in an excel file that gives an overview of the métiers allocated.

5 R script

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

5.1 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
- lubridate

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

5.2 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	gear_FR	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	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	GNS	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	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	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	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	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	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	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	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	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	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	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	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	OTB	115	NA	NA	YYY	NA	weight	20.00	100.0000

5.3 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)*.

```r
# Prepare input data

input.data[, ':='(EUR, as.numeric(EUR))]

input.data[, ':='(KG, as.numeric(KG))]

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.

```r
# Assign RCG names to the input data
input.data <- merge(input.data, area.list, all.x = T, by = "area")
```

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.

```
# Assign species category to the input data
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.

```
# Process input data In the variable called sequence.def
# please include all columns that will constitute a fishing
# sequence This variable will be used as a key for grouping
# operations
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.

```r
# Select a measure to determine the dominant group at a
# sequence level. If at least one species in a sequence has
# 'value' in a measure column then all species in that
# sequence get the same measure.
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.

```r
# Determine the dominant group for each 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.

```
# Apply DWS rules
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.

```
# Assign metier level 6
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

NEED TO ADD TEXT HERE

```
# Missing metiers
input.data[, ':='(month = month(dmy(fishing_day)), quarter = quarter(dmy(fishing_day)))]
input.data[, ':='(vessel_length_group, cut(vessel_length, breaks = c(0,
    10, 12, 18, 24, 40, Inf), right = F))]
input.data.sequances <- unique(input.data[metier_level_6 != "MIS_MIS_0_0_0",
    .SD, .SDcols = c(sequence.def, "seq_dom_group", "metier_level_6",
        "gear_FR", "month", "quarter", "vessel_length_group")])

input.data[metier_level_6 == "MIS_MIS_0_0_0", ':='(metier_level_6,
    as.character(pmap(list(vessel_id, month, area, seq_dom_group,
        quarter, year, vessel_length_group, gear_FR), function(v,
        m, a, sg, q, y, vg, gfr) getMissingMetier(v, m, a, sg,
        q, y, vg, gfr))))]
```

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.

```
# Analyze vessel patterns
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"))
```
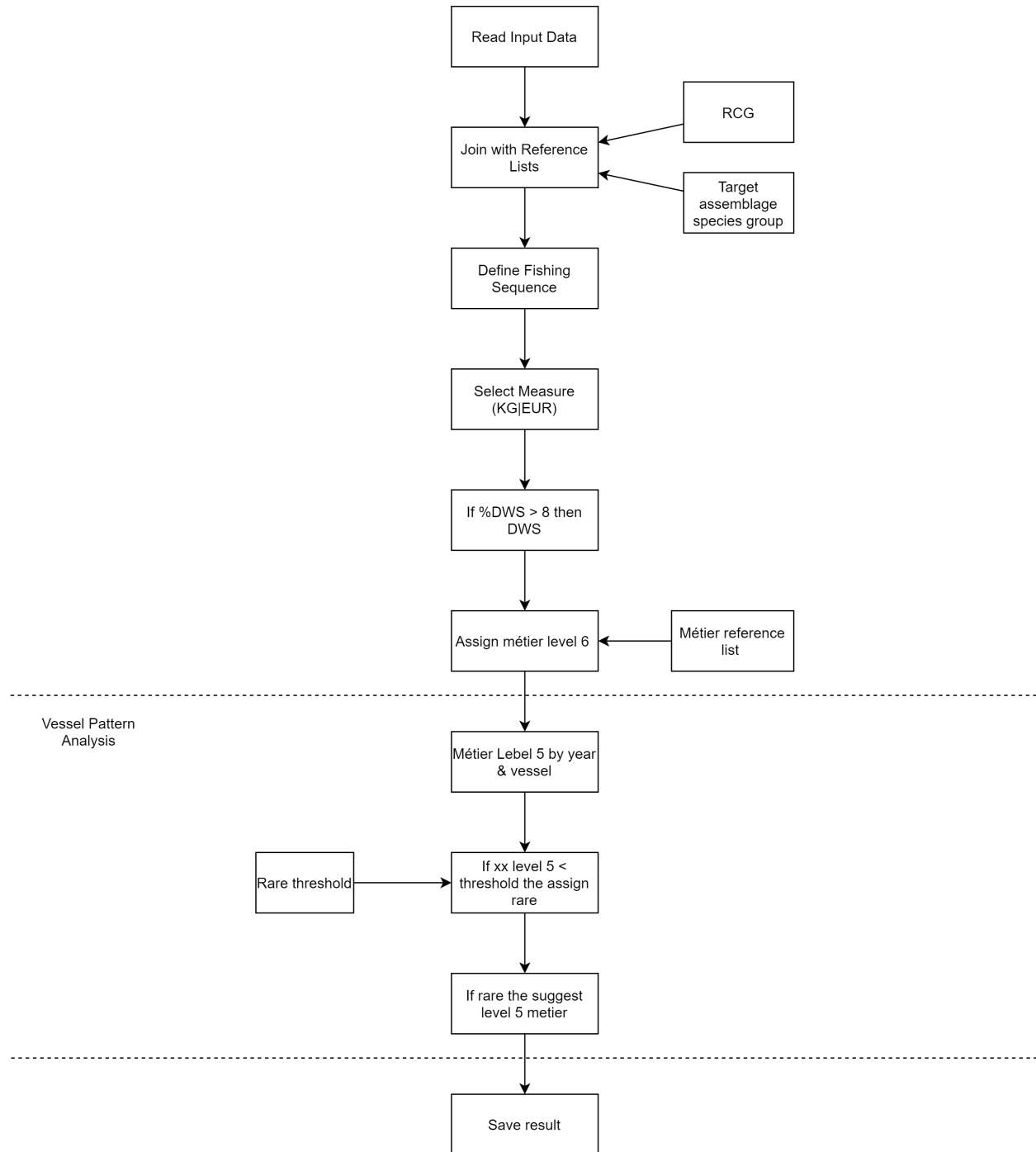
year	vessel_id	metier_level_5	seq_no_lvl5	seq_perc_lvl5
2017	GDY-555	OTM_SPF	1	100.00000
2018	BBB-3	GNS_DEF	1	100.00000
2018	AAA-1	OTB_DEF	2	66.66667
2018	ZZZ-2	OTB_DEF	1	100.00000
2010	ZZZ-2	OTB_DEF	1	100.00000
2020	DDD-4	GNS_FWS	1	11.11111
2020	DDD-4	GNS_DEF	1	11.11111
2020	DDD-4	OTB_DEF	5	55.55556
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_".

```r
# Specify the percentage threshold of the number of sequences
# below which a metier will be considered rare
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

6 Workflow diagramm

```
                    ┌─────────────────┐
                    │ Read Input Data │
                    └────────┬────────┘
                             │
                             ▼                      ┌─────────────┐
                    ┌─────────────────┐◄────────────│     RCG     │
                    │ Join with       │             └─────────────┘
                    │ Reference Lists │             ┌─────────────┐
                    └────────┬────────┘◄────────────│   Target    │
                             │                      │ assemblage  │
                             ▼                      │species group│
                    ┌─────────────────┐             └─────────────┘
                    │ Define Fishing  │
                    │ Sequence        │
                    └────────┬────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │ Select Measure  │
                    │ (KG|EUR)        │
                    └────────┬────────┘
                             │
                             ▼
                    ┌─────────────────┐
                    │ If %DWS > 8 then │
                    │ DWS             │
                    └────────┬────────┘
                             │
                             ▼                      ┌─────────────┐
                    ┌─────────────────┐◄────────────│   Métier    │
                    │Assign métier    │             │ reference   │
                    │level 6          │             │    list     │
                    └────────┬────────┘             └─────────────┘
```

Vessel Pattern
Analysis

```
                             ▼
                    ┌─────────────────┐
                    │ Métier Lebel 5  │
                    │ by year & vessel│
                    └────────┬────────┘
                             │
   ┌──────────────┐          ▼
   │Rare threshold│─────►┌─────────────────┐
   └──────────────┘      │ If xx level 5 < │
                         │ threshold the   │
                         │ assign rare     │
                         └────────┬────────┘
                                  │
                                  ▼
                         ┌─────────────────┐
                         │If rare the      │
                         │suggest level 5  │
                         │metier           │
                         └────────┬────────┘
```

```
                                  ▼
                         ┌─────────────────┐
                         │  Save result    │
                         └─────────────────┘
```
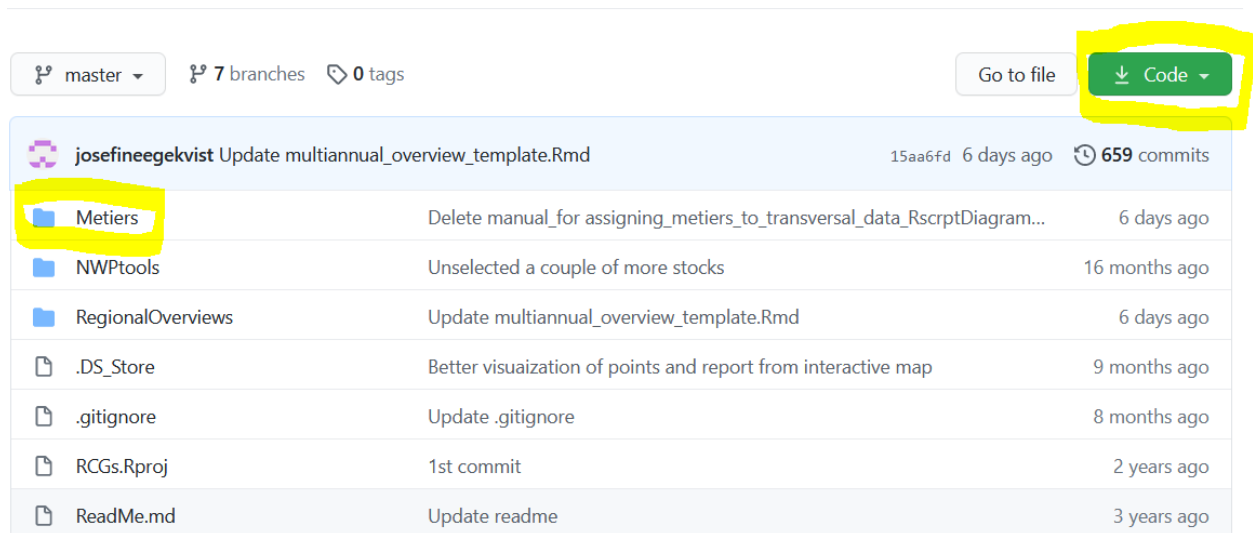
Annex 1: Getting the code from GitHub

There are several ways to get the code from GitHub, three different ways are described here:

Download zip-file directly from GitHub

1. On the GitHub link: https://github.com/ices-eg/RCGs there is a green button called "Code". If clicking on that there is a possibility to download Zip. This will download all content of RCG ISSG's at GitHub.

2. Unzip the file at your computer. Under the "Metiers" -> "Scripts" folders, you can find the script_metiers_test.R file. Under the "Functions" folder the functions that the script reads are located. In the beginning of the script, make sure that it points to the Functions folder to read the functions.

3. To open the project from your computer under the "Metiers" folder you should run "Metiers.Rproj". Then the path to all functions will be valid.


Using GitHub Desktop

1. Download GitHub Desktop here: https://desktop.github.com/

2. Define a folder where a copy of the GitHub content should be copied to

3. Set up the connection between GitHub RCGs url: https://github.com/ices-eg/RCGs and the folder on your computer.

Using Git for Windows

1. Install Git for Windows. You can download git from here https://git-scm.com/download/win. During the installation process you don't have to change any options.

2. Once you have git installed you can start your RStudio. In the top right corner there is a button called "Project". You should click it and select New Project.
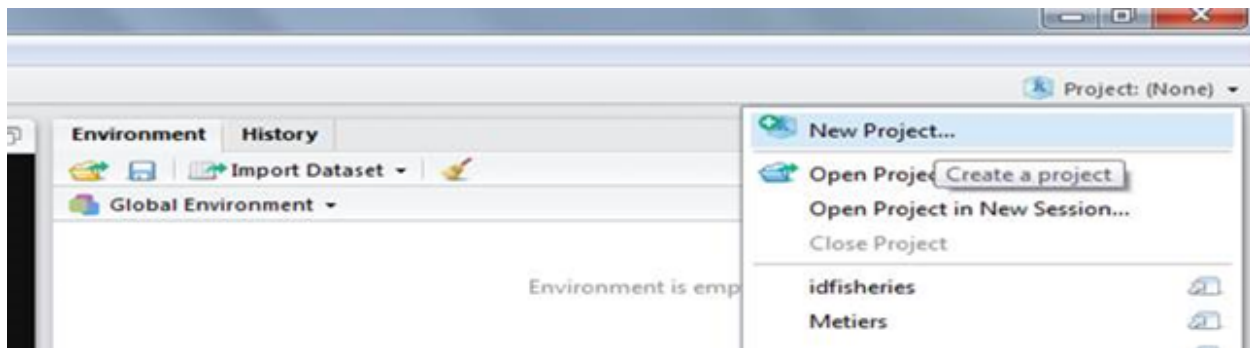
3. A new window will show up. On that window select "Version Control". Next select "Git". In the next window you should enter the repository URL which is https://github.com/ices-eg/RCGs.git Note that RStudio for Windows prefers for Git to be installed below C:/Program Files and this appears to be the default. This implies, for example, that the Git executable on Windows system should be found at C:/Program Files/Git/bin/git.exe. In case when Git was installed in another directory you should select Tools -> Global options -> Git/SVN, and specify a path to the Git executable on your computer:
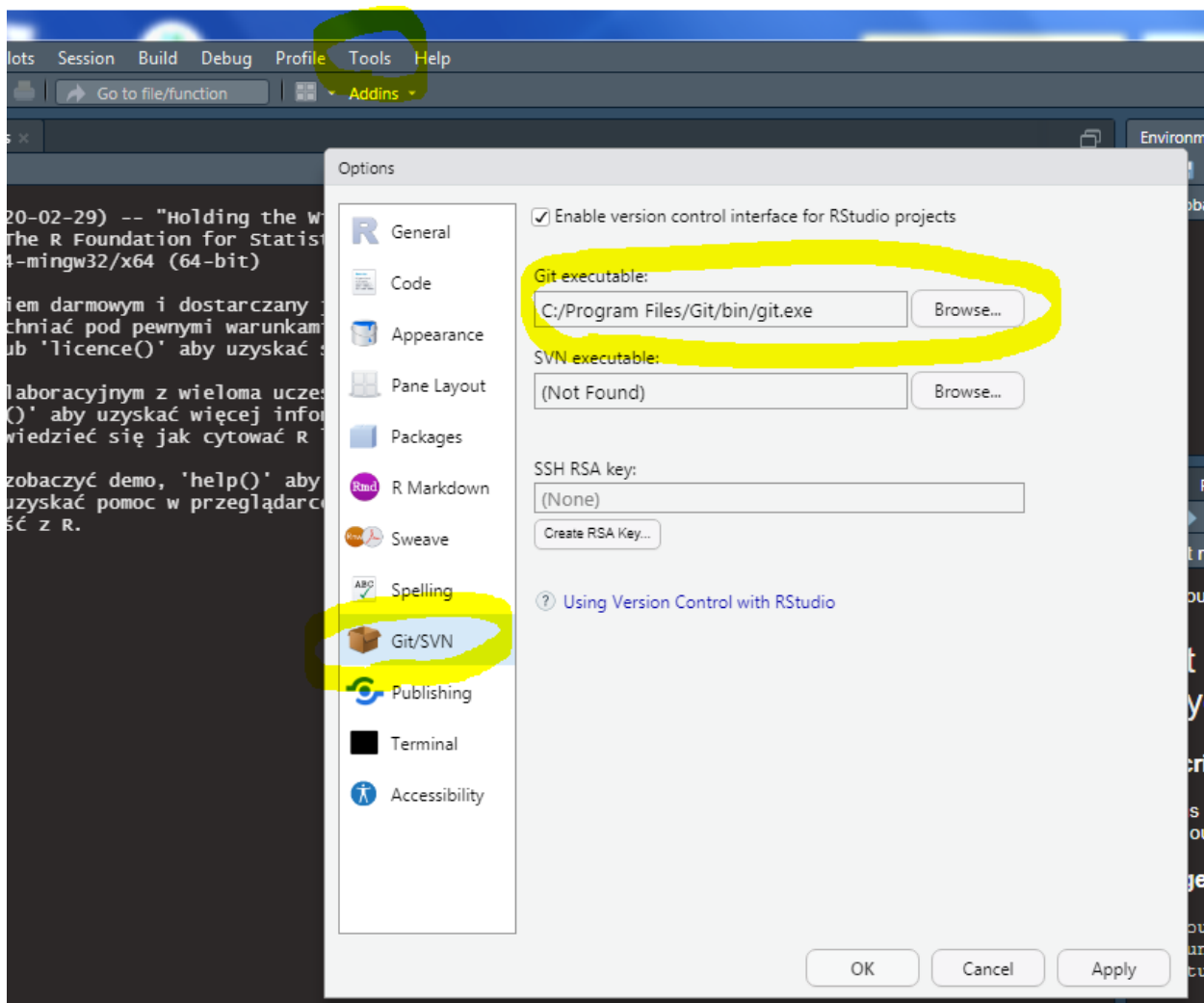
4. In the field "Create project as a subdirectory of" you can choose the directory on your hard disk where you want this project to be downloaded. And then click "Create project". After a while RStudio will download (clone) the repository from github to your disk. This repository contains data from many RCG sobgroups. The content for the metier assignment is in the "Metier" subfolder.

Contact Information

*Contact persons:

Questions regarding code lists: Josefine Egekvist (jsv@aqua.dtu.dk)

Questions regarding the R-script: Maciej Adamowicz (madamowicz@mir.gdynia.pl)