

Rmd files for RCG data calls

Second Training on quality checks
17/04/2024



COISPA

Isabella Bitetto,
Water Zupa,
Loredana Casciaro,
Cosmidano Neglia



COISPA



CNR
IRBIM
Istituto per lo studio e la
ricerca in biologia marina
e in ecologia



RCG data call formats

The RCG MED & BS data call requires 2 tables:

- 1) Detailed data (starting from the trip (eventually by haul), to the individual length measurement, individual weight, age, maturity, etc...) -> **CS table**
- 2) Aggregated landing data -> **CL table**

The template is on the sharepoint:

[https://coispa.sharepoint.com/:x:/s/1stQualiTraintrainingondat
aqualitychecks/ESMu2yjR1sJMqeNIL17IH84BWRoMz7Z_XjgxIk
QidsJrag?e=ddhDKI](https://coispa.sharepoint.com/:x:/s/1stQualiTraintrainingondat
aqualitychecks/ESMu2yjR1sJMqeNIL17IH84BWRoMz7Z_XjgxIk
QidsJrag?e=ddhDKI)

The Excel file contains codification for each field to be filled in.

CL table: landing

```
> head(CL)
  flag_country year quarter month  area species
1          ITA 2008         1     1 GSA18 Parapenaeus longirostris
2          ITA 2008         1     2 GSA18 Parapenaeus longirostris
3          ITA 2008         1     3 GSA18 Parapenaeus longirostris
4          ITA 2008         2     4 GSA18 Parapenaeus longirostris
5          ITA 2008         2     5 GSA18 Parapenaeus longirostris
6          ITA 2008         2     6 GSA18 Parapenaeus longirostris
 fishing_activity_category_national fishing_activity_category_eu_16
1                                NK OTB_DEF_>=40_0_0
2                                NK OTB_DEF_>=40_0_0
3                                NK OTB_DEF_>=40_0_0
4                                NK OTB_DEF_>=40_0_0
5                                NK OTB_DEF_>=40_0_0
6                                NK OTB_DEF_>=40_0_0
harbour official landings weight official landings value
1      NK                                40131.53          305405.2
2      NK                                59446.60          453739.8
3      NK                                45953.19          338358.1
4      NK                                63345.98          452249.8
5      NK                                89739.98          619798.9
6      NK                                32244.76          244364.5
```

Main data in CL

- official_landings_weight
- official_landings_value

CS table: sampling

1	sampling_type	21	area
2	flag_country	22	fishing_activity_category_national
3	year	23	fishing_activity_category_eu_l6
4	trip_code	24	species
5	harbour	25	catch_category
6	n_sets_hauls	26	weight
7	days_at_sea	27	subsample_weight
8	sampling_method	28	sex
9	aggregation_level	29	maturity_method
10	station_number	30	maturity_scale
11	duration_fishing_operation	31	maturity_stage
12	initial_latitude	32	ageing_method
13	initial_longitude	33	age
14	final_latitude	34	length_code
15	final_longitude	35	length_class
16	depth_fishing_operation	36	number_at_length
17	water_depth	37	commercial_size_category_scale
18	catch_registration	38	commercial_size_category
19	species_registration	39	fish_id
20	date	40	individual_weight

Main data in CL

- Harbour
- Fishing_activity_category_national
- Fishing_activity_category_European_lvl_6
- Sampling_method
- Trip_code
- Initial_latitude
- Initial_longitude
- Final_latitude
- Final_longitude
- Number_at_length
- Individual_weight
- Species
- Maturity_Stage
- Sex
- Age
- Fish_ID
- Length_class
- Trip_code
- Commercial_size_category

RDBqc: data quality checks for RCG format (CS and CL)

Consistency (on CS) of:

Table	Check function
CL	check_CL
CS	check_AL
CS	check_LFD
CS	check_LFD_comm_cat
CS	check_lw
CS	check_mat
CS	check_mat_ogive
CS	summarize_trips
CS	summarize_ind_meas
CS	check_loc

1. LFD by year according to commercial category;
2. length-weight data;
3. age-length data;
4. sex-length-maturity stage;
5. Summary of the number of individual biological data (length, sex, maturity, weight and age) collected by trip;
6. Summary of the number of trips/hauls monitored by year, port, métier, sampling method;
7. Mapping of the hauls position or the average trip position by year.

Checks on CL:

- temporal coverage of landing and landing value data (according to different timeframes: month, quarter, year),
- spatial coverage (GSA and harbour) by year,
- coverage of species and métier, qualitative checks of time series.

```
<!-- USER DEFINED SOURCE FILES -->

```{r::setup2, include=FALSE}
reading files
UNCOMMENT THE FOLLOWING LINES FOR STAN-ALONE USE
SELECT THE APPROPRIATE FILE PATHS ON THE LOCAL FOLDERS
CL <- read.csv("D:/OneDrive -- Coispa Tecnologia & Ricerca S.C.A.R.L/RDB/Workshop
2/dati/RCG/dc_rcg_landings_ITA_SP.csv", sep=";")
CS <- read.csv("D:/OneDrive -- Coispa Tecnologia & Ricerca S.C.A.R.L/RDB/Workshop
2/dati/RCG/dc_rcg_sampling_ITA_SP.csv", sep=";")

SPs <- c("Merluccius merluccius", "Mullus barbatus", "Parapenaeus longirostris")
MS <- "ITA"
GSAs <- c("GSA18")
stages_for_immatures <- c("0", "1", "2a")
```
```



Used in «RCG_check_mat_ogive» function

In case it is commented in the Rmd, the default value is c("0", "1", "2a")

Practical session

On the sharepoint there are RCG CS and CL tables with errors to be used in the practical session:

https://coispa.sharepoint.com/:f:/s/1stQualiTraintrainingondataqualitychecks/Emmudg5qvS1Nm_8QhEHAY3UB8DKnI0IkTrJL4qFpu4bthw?e=Tz8j8v