Second QualiTrain training on quality checks

GFCM DCRF data call formats

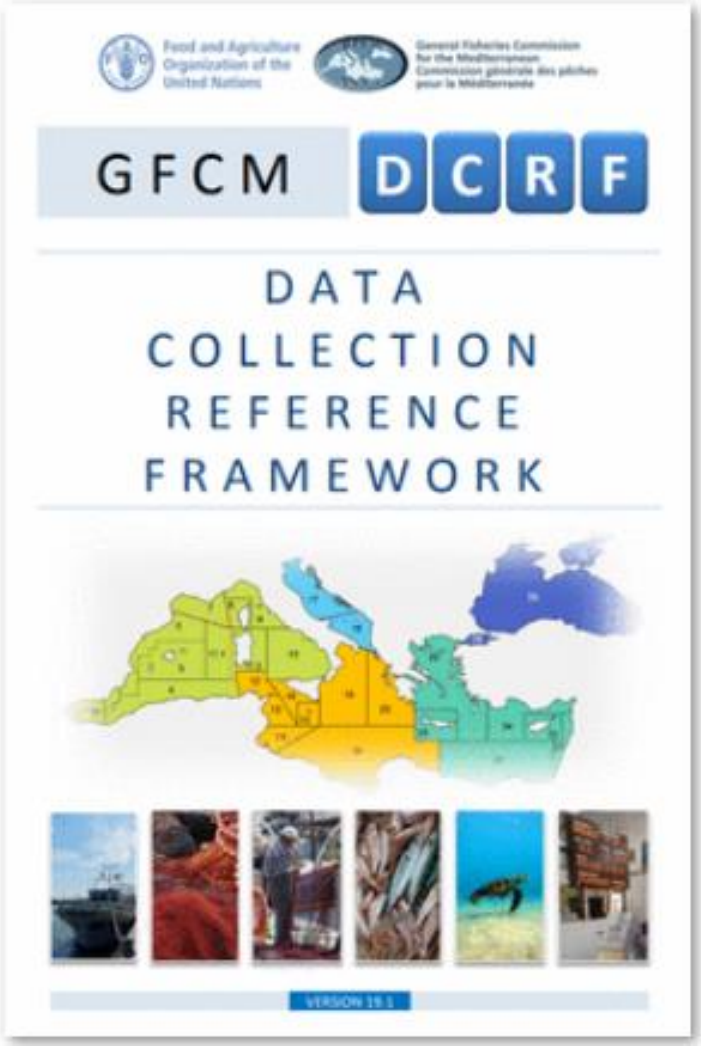
16th April 2024

Isabella Bitetto &
Walter Zupa

Ad hoc R functions have been designed for specific **consistency quality checks** for the GFCM DCRF tasks on biological information stored in:

TASK II.2 - CATCH DATA PER SPECIES
TASK III - INCIDENTAL CATCH
TASK VII.2 - LENGTH DATA
TASK VII.3.1 - SIZE AT FIRST MATURITY
TASK VII.3.2 - MATURITY DATA

https://www.fao.org/gfcm/data/dcrf/es/



RDBqc: data quality checks for GFCM DCRF data call

Table	Check functions
task 2.2	GFCM_Check_empty_field_II2.R
task 2.32	GFCM_Check_record_duplicated_II2.R
task 2.2	GFCM_cov_task2.2
task 3	GFCM_Check_empty_field_III.R
task 3	GFCM_Check_record_duplicated_III.R
task 3	GFCM_cov_task3
task 7.2	GFCM_Check_empty_field_VII2.R
task 7.2	GFCM_Check_record_duplicated_VII2.R
task 7.2	GFCM_relatioship_length_weight_VII2.R
task 7.2	GFCM_Check_lfd_VII2.R
task 7.3.1	GFCM_Check_empty_field_VII31.R
task 7.3.1	GFCM_Check_record_duplicated_VII31.R
task 7.3.1	GFCM_Check_L50_VII31.R
task 7.3.2	GFCM_Check_empty_field_VII32.R
task 7.3.2	GFCM_Check_record_duplicated_VII32.R
task 7.3.2	GFCM_check_species_catfau_VII32.R
task 7.3.2	GFCM_relatioship_length_weight_VII32.R

- Duplicated records;
- Empty fields,
- Consistency of length information (min-max),
- Length-weight,
- Maturity.

TASK II.2 - CATCH DATA PER SPECIES

	A	B	C	D	E	F	G	H	I	J
1	II.2 - Catch data per species									
2	country	reference_year	ographical subarea (GS	Fleet segment	Species(scientific name or 3-alpha code)	Total landing (in tons) per species	Total discards (in tons) per species (if any)	Total catch (in tons) per species	Comments (if any)	
3	ITA	2019	18	L-05	DPS	8.79	NA	8.79	Discard derogation following the results	
4	ITA	2019	18	T-10	DPS	5.63	0.24	5.87	Discard volume by fleet segment obtaine	
5	ITA	2019	18	T-11	DPS	1988.99	85.97	2074.96	Discard volume by fleet segment obtaine	
6	ITA	2019	18	T-12	DPS	182.52	7.89	190.41	Discard volume by fleet segment obtaine	
7	ITA	2019	18	L-05	MUT	0.66	NA	0.66	Discard derogation following the results	
8	ITA	2019	18	P-05	MUT	29.72	NA	29.72	Discard derogation following the results	
9	ITA	2019	18	P-06	MUT	28.31	NA	28.31	Discard derogation following the results	
10	ITA	2019	18	P-07	MUT	2.31	NA	2.31	Discard derogation following the results	
11	ITA	2019	18	T-10	MUT	8.45	2.35	10.8	Discard volume by fleet segment obtaine	
12	ITA	2019	18	T-11	MUT	688.15	191.11	879.26	Discard volume by fleet segment obtaine	

In this table the information on the landing and discard by species and GFCM fleet segment is stored.
https://www.fao.org/gfcm/activities/fisheries/former-fisheries-pages/fisheriescopy/fleet/segmentation/en/

Be aware that the GFCM fleet segments are not the same of DCF !

The fleet segments relevant in each GSA are provided by the GFCM in the template provided to NC (to be filled in); previously, the NC provided a feedback to GFCM on the appropriateness of the fleet segments list by GSA.

The MS should provided consistent data by species among the two fleet segmentation.

Vessel groups			Length classes (LOA)			
			< 6 m	6 - 12 m	12-24 m	> 24 m
Polyvalent	P	Small-scale vessels without engine using passive gears	P-01	P-02	P-03	P-04
			P-13			
		Small-scale vessels with engine using passive gears	P-05	P-06	P-07	P-08
		Polyvalent vessels	P-09	P-10	P-11	P-12
					P-14	
Seiners	S	Purse seiners	S-01	S-02	S-03	S-04
					S-09	
		Tuna seiners	S-05	S-06	S-07	S-08
					S-10	
Dredgers	D	Dredgers	D-01	D-02	D-03	D-04
				D-05		
Trawlers	T	Beam trawlers	T-01	T-02	T-03	T-04
		Pelagic trawlers	T-05	T-06	T-07	T-08
				T-13		
		Trawlers	T-09	T-10	T-11	T-12
Longliners	L	Longliners	L-01	L-02	L-03	L-04
				L-05		

One way to do it is…

To define (internally to your institute or at Country level) a communication table between GFCM fleet segments and metier-LOA defined within DCF to provide the sampled and raised data consistently with the DG MARE Med&BS Data Call.

	Fleet_segment	Description	METIER	LOA	GEAR_ACatch	MESH_SIZE_RANGE_ACatch	FISHERY_ACatch
1	D-01	Dredgers (< 6)	DRB_MOL_0_0_0	VL0006	DRB	NK	MOL
2	D-02	Dredgers (6 - 12)	DRB_MOL_0_0_0	VL0612	DRB	NK	MOL
3	D-03	Dredgers (12 - 24)	DRB_MOL_0_0_0	VL1218	DRB	NK	MOL
4	D-03	Dredgers (12 - 24)	DRB_MOL_0_0_0	VL1824	DRB	NK	MOL
5	D-04	Dredgers (> 24)	DRB_MOL_0_0_0	VL2440	DRB	NK	MOL
6	D-04	Dredgers (> 24)	DRB_MOL_0_0_0	VL40XX	DRB	NK	MOL
7	L-01	Longliners (< 6)	LLS_DEF_0_0_0	VL0006	LLS	NK	DEF
8	L-02	Longliners (6 - 12)	LLS_DEF_0_0_0	VL0612	LLS	NK	DEF
9	L-03	Longliners (12 - 24)	LLS_DEF_0_0_0	VL1218	LLS	NK	DEF
10	L-03	Longliners (12 - 24)	LLS_DEF_0_0_0	VL1824	LLS	NK	DEF

TASK III - INCIDENTAL CATCH

country	reference_year	Date (dd/mm/yyyy)	Geographical subarea (GSA)	Source of data	Fleet segment	Fishing gear (if available)	Group of vulnerable species	Family (if available)	Species - Scientific name (if available)	Total number of individuals caught	Total weight of individuals caught (if available)	Number of individuals released alive (if available)	Number of dead individuals (if available)	Number of individuals released in unknown status (if available)
ITA	2020	44142	18	SU	T-12	TB	Sharks, Rays,	Myliobatidae	Pteromylaeus bo	2	0.0408	2	0	0
ITA	2020	44142	18	SU	T-12	TB	Sharks, Rays,	Dasyatidae	Dasyatis pastinac	1	0.00055	1	0	0
ITA	2020	44149	18	SU	T-12	TB	Sharks, Rays,	Rajidae	Raja polystigma	2	0.00078	0	2	0
ITA	2020	44140	18	SU	T-12	TB	Sharks, Rays,	Rajidae	Leucoraja circular	1	0.00036	0	1	0
ITA	2020	43896	18	BS	T-11	TB	Sharks, Rays,	Dasyatidae	Dasyatis pastinac	3	0.00153	0	3	0
ITA	2020	43921	18	BS	P-06	GTR	Sharks, Rays,	Dasyatidae	Dasyatis pastinac	2	0.00072	0	2	0
ITA	2020	43954	18	BS	L-05	LLS	Sharks, Rays,	Myliobatidae	Pteromylaeus bo	10	0.1327	0	10	0

In this table the information on incidental catch is stored, with detail on the number of individuals released alive, dead and the weight.

TASK VII.2 - LENGTH DATA

- In Table VII.2 the sampled and the expanded LFDs by species and by fleet segment are stored.
- Only the commercial fraction should be reported.

country	reference_ year	Geographic al subarea (GSA)	Source of data	Name of the scientific survey (if any)	Fleet segment (only if from fishery-dependent data)	Species (scientific name or 3-alpha code)	Length unit	Length	Number of individuals measured (per length classes)	Weight of individuals sampled per length classes (kg)	Number of individuals expanded (per length classes)	Comments (if any)
ITA	2019	18	BS		P-07	MUT	cm	10.5	1	0.01341891	400	
ITA	2019	18	BS		P-07	MUT	cm	11	1	0.01546473	800	
ITA	2019	18	BS		P-07	MUT	cm	11.5	1	0.01791485	3601	
ITA	2019	18	BS		P-07	MUT	cm	12	1	0.02027359	6401	
ITA	2019	18	BS		P-07	MUT	cm	12.5	2	0.022904	8402	
ITA	2019	18	BS		P-07	MUT	cm	13	2	0.02583199	9202	
ITA	2019	18	BS		P-07	MUT	cm	13.5	2	0.02903021	8602	
ITA	2019	18	BS		P-07	MUT	cm	14	2	0.03247698	7602	
ITA	2019	18	BS		P-07	MUT	cm	14.5	2	0.03613063	7802	
ITA	2019	18	BS		P-07	MUT	cm	15	1	0.04013993	3601	
ITA	2019	18	BS		P-07	MUT	cm	15.5	1	0.04431141	5801	

Task VII.2

FIELDS	DEFINITION OF VARIABLES	PK*
Country	*The ISO 3-alpha code of the country under which all vessels operate (refer to the FAO country codes list)*	x
Reference year	*The year the collected data refer to*	x
GSA	*The geographical subarea code as reported in Appendix J of this manual (Res. GFCM/33/2009/2)*	x
Source of data	*Source of data: **BS** (biological sampling on board, at landing place, at market, self-sampling, others) or **SU** (scientific survey)*	x
Name of the scientific survey* **	*Full name of the scientific survey. If more than one survey is in place, this line should be repeated*	x
Fleet segment**	*The fleet segment code (Appendix B of this manual) or the code corresponding to the merged fleet segments (see section 6.2)*	x
Species	*The 3-alpha code identifying the species (Appendix A of this manual)*	x
Length	*The value of the length class of the measured individuals. Length classes should be reported in centimetres (cm), as a whole number, or in half centimetres (e.g. 0.5, 1.0, 1.5 cm, etc.) for fish (including elasmobranchs) and cephalopods. For crustaceans, length classes should be reported in millimetres (e.g. 1, 2, 3, 4 mm, etc.) (see section 6.3.1)*	x
Number of individuals sampled (per length classes)	*The total number (n) of individuals measured for that fleet segment and length class for the reference year*	
Weight of individuals sampled (per length classes)	*Mean individual weight (kg) by length class for that fleet segment and for the reference year*	
Number of individuals expanded (per length classes)	*The total number of individuals, for a given length class, expanded to the total production (e.g. landing) of each fleet segment. In case of an experimental survey, it should be expanded to the total catch of that survey*	

*** only if from fishery-dependent data*

**** if any*

**data fields composing the primary key (PK) of the table*

TASK VII.3.1 - SIZE AT FIRST MATURITY

Table VII.3.1 contains information of the size at first maturity by species, by sex, by year.

country	reference_year	Geographical subarea (GSA)	Species (scientific name or 3-alpha code)	Sex	L50	Reference (if any)	Comments (if any)
ITA	2019	18	DPS	F	17.7355		
ITA	2019	18	DPS	M	11.9061		
ITA	2019	18	MUT	F	11.8584		
ITA	2019	18	MUT	M	10.4308		
ITA	2020	18	DPS	F	17.91		
ITA	2020	18	DPS	M	11.803		
ITA	2020	18	MUT	F	11.3767		
ITA	2020	18	MUT	M	10.3812		
ITA	2021	18	DPS	F	17.9846		
ITA	2021	18	DPS	M	12.3246		
ITA	2021	18	MUT	F	11.3449		
ITA	2021	18	MUT	M	10.0787		

TASK VII.3.2 - MATURITY DATA

- In Table VII.3.2 the maturity data (sampled and expanded) are provided by species, se, year and fleet segment.
- Same consideration on GFCM fleet segments of Table II.2, VII.2 and III.

country	reference_year	GSA	Source of data	Name of the scientific survey	Fleet segment	Species	Length_unit	Length	Sex	Maturity	Number of individuals sampled per length class sex	Weight of individuals sampled per length class sex	Number of individuals expanded per length class sex
ITA	2019	18	BS		T-10	DPS	mm	10	F	Crustacear	1	0.000996016	27
ITA	2019	18	BS		T-10	DPS	mm	11	F	Crustacear	1	0.001311431	89
ITA	2019	18	BS		T-10	DPS	mm	12	F	Crustacear	1	0.001605501	106
ITA	2019	18	BS		T-10	DPS	mm	13	F	Crustacear	1	0.001935129	310
ITA	2019	18	BS		T-10	DPS	mm	14	F	Crustacear	1	0.002301502	696
ITA	2019	18	BS		T-10	DPS	mm	15	F	Crustacear	1	0.002705758	5699
ITA	2019	18	BS		T-10	DPS	mm	16	F	Crustacear	1	0.003148989	19833
ITA	2019	18	BS		T-10	DPS	mm	17	F	Crustacear	1	0.003632249	3391

Technical steps

A Vignette has been developed to show the functioning of each function with a dummy dataset and is available on GitHub repository:

https://github.com/COISPA/RDBqc/tree/main/RDBqc



GFCM data format

Task II.2 table

Check empty fields in GFCM Task II.2 table

The function `check_EF_taskII2` checks the presence of not allowed empty data in the given table, according to the GFCM Data Collection Reference Framework (DCRF, V. 20.1). The function returns two lists. The first list gives the number of NA for each reference column.

```
check_EF_taskII2(task_ii2, verbose=FALSE)[[1]]
#> Reference_Year            CPC            GSA        Segment        Species
#>              0              0              0              0              0
#>        Landing          Catch
#>              0              0
```

The second list returns the index of each NA in the reference column.

```
check_EF_taskII2(task_ii2, verbose=FALSE)[[2]]
#> $Reference_Year
#> integer(0)
#>
#> $CPC
#> integer(0)
#>
```