Documentation of the Regional Database and Estimation System Data Model

RDBES Data Model doc. v. 1.18

12 November 2020

Table of Contents

Philosophy of the data model

The data model that is described in this report is intended to be used in the new Regional Database and Estimation System (RDBES) that is currently being developed by ICES. Some more details about the RDBES can be found in Annex II.

The aims of the RDBES are:

1) To ensure that data can be made available for the coordination of regional fisheries data sampling plans, in particular for the EU DC-MAP Regional Coordination Groups (RCGs),

2) To provide a regional estimation system such that statistical estimates of quantities of interest can be produced from sample data,

3) To increase the data quality, documentation of data and ensuring of approved estimation methods are used,

4) To serve and facilitate the production of fisheries management advice and status reports,

5) To increase the awareness of fisheries data collected by the users of the RDBES and the overall usage of these data.

In the longer term it is expected that the new RDBES will replace both the current RDB and InterCatch systems and provide a single platform for countries to produce statistical estimates of quantities of interest (such as discards and age or length distributions) which will then be used as inputs for assessment working groups.

The RDBES data model is an evolution of the work already done in defining and using the current Regional Database (RDB) and the COST data models and functions. The current RDB data model provides a common data structure to describe commercial sampling data at a disaggregated level, and commercial landings and effort data at an aggregated level. **The significant difference in the new RDBES model is that it provides a common structure to describe both the disaggregated sampling data and, most importantly, how it was sampled.**

The RDBES data model should allow a variety of different estimation techniques to be used including the COST functions that make use of Age-Length keys, ratio estimators, and unbiased designed based estimations. Future workshops will be held to further test and develop how these estimation techniques can be used with the new model and the model adjusted if needed.

The RDBES data model should be seen as part of the movements towards:

- Statistically Sound Sampling Schemes (4S),
- Greater regional coordination,
- Transparent Assessment Framework (TAF),
- Improved estimates to ICES assessments.

Whilst the RDBES data model is designed to hold 4S data it will also be able to store data that is not sampled in a statistical manner – this is important so that data from current non- statistical programmes can be uploaded and historical data can also be stored. However, in doing so, the new RDBES data model flags those data thus allowing their appropriate interpretation during estimation.

Key Concepts in RDBES Data Model

The most important step in populating the RDBES data model is that you must first specify the design of your sampling programmes. This is because, contrary to the RDB, there is not a single set of tables that everyone must populate and upload: With a few exceptions, the tables that need to be populated and uploaded depend on the specific sampling design(s) each country used.

Specifying the sampling design is not always clear-cut, particularly in the case of old undocumented data or sampling schemes that depart greatly from the principles of probability-based sampling. Useful hands-on guidance on the documentation of sampling plans and the identification of sampling designs (including the definition of the hierarchies involved in multi-stage sampling) can be found in ICES EGs reports such as WKPICS, SGPIDS, WKRDB 2014-1 and WGCATCH.

The data model for the RDBES is shown in the spreadsheet "RDBES Data Model.xlsx". The first sheet is an overview of all the tables, and in the following sheets each table is defined in details. Depending on the sampling scheme a combination of tables will have to be populated with data.

Sampling Data (CS)

To allow for the variety of sampling designs used in different institutes and countries the RDBES uses a slightly different approach to most database designs you might be used to. Rather than specifying tables and the fixed relationships between them the RDBES identifies a number of different **sampling hierarchies** – these represent the different hierarchical sampling techniques that are used in practice. Two types of hierarchy are used in the model – the **upper hierarchy** describes how a sample is selected, whilst the **lower hierarchy** describes what type of length-frequency or biological variables are measured for that sample. The sampling hierarchies are defined in terms of the **tables** they consist of and the **relationships** between them. All tables contain **columns**. Some of these columns are described as **design variables** (DV) and represent important sampling concepts like stratification, population and sample size, inclusion probability or sampling method. These will be described in more details in Section "The new design variables". Tables will typically appear in a number of different hierarchies. There are two types of tables: **hierarchy tables** and **auxiliary tables**. Hierarchy tables are those which contribute to the selection of a sample, whilst auxiliary tables contain data of interest which does not directly contribute to the sampling design. A table can appear as a hierarchy table in one sampling hierarchy whilst it can appear as an auxiliary table in another. In most hierarchy tables the **rows** represent individual samples taken at that particular stage of the sampling hierarchy.

To populate the data model:

1) Identify the different sampling schemes you will use as the source data.
2) Ensure you understand how the sampling scheme is defined.
3) For each sampling scheme you have identified in point (1) identify which **upper hierarchy** is the best match by using the spreadsheet "Hierarchies and relating tables" to see the situations under each hierarchy and read the "Description of each hierarchy" in the Annex I in this document, to determine which hierarchy to use.
4) Extract your national data to populate the **upper hierarchy** tables. Refer to section "The new design variables" for details about how to complete the new **design variables.**
5) Now identify which **auxiliary tables** you will be supplying data for and populate them.
6) For each sample you will need to decide which **lower hierarchy** you will be using – this describes whether the sample contains length-frequency data, biological measurements, or both or none.
7) The tables of the **lower hierarchy** can now be populated.
8) Data for each hierarchy have to be in a separate data files one for each hierarchy e.g. 2, 7, 8

Aggregated Commercial Landings (CL) and Commercial Effort (CE)

In the RDBES data model the aggregated population data are generally used, e.g., in analyses of landings and effort at marine region level, is kept in a commercial landings (CL) and commercial effort (CE) table. These tables are identical to the ones presently available in the RDB. Their data model is available on the ICES website[1].

These data formats are collecting aggregated data from national commercial fisheries. It allows for both official and scientific estimates of landings and effort. The RDBES should make it possible to use the data to populate the future ICES data call and preferably also the STECF FDI data call

Below is a description on how to fill in fields in the two tables, where it is not obvious.

Fields in Commercial Landings (CL) data format

Data sources: In the dataset, it is possible to give the official weight and a scientific weight. The scientific weight could e.g. be estimates for the small-scale fisheries based on sampling. To make it transparent, it has been made possible to add information on data sources in the data format. Data should not be duplicated, meaning that there should only be given one estimate on the scientific weight, which can be based either on official data or sampling data. The data source for the landings value is also included in the data format. The data sources are given in four fields:

- *Data type of scientific weight*: The scientific weight can be based on either official data or data based on estimates. This field can be filled in with "Official data if data are based on official data like logbooks, sales notes or other declarative forms or "Estimate" if it is based on sampled data.
- *Data source of scientific weight*: The data source of the scientific weight can be set to "Logbook", "Sales notes", Other declarative forms", "Combination of official data" or "Sampling data".
- *Sampling scheme behind scientific weight*: If sampling data are used to estimate the scientific weight, it should preferable be specified from the *Sampling scheme*. The national options of sampling from the *Sampling scheme* are maintained in a code list in the RDBES.
- *Data source landings value*: The value of the landings (Euro) can be obtained from different sources: "Sales notes" if they are derived directly from sales notes", "Average prices", "Combination of sales notes and average prices" or "Other" if the value of the landings are estimated from other sources.

Landing category and Catch category: The landing category indicates the intended usage of the landing while the catch category indicates the possible fractions of the catch:

- *Landing category*: Can be set to Industry (Ind), human consumption (HuC) or None for logbook registered discards.
- *Catch category*: This field should match the Catch category field from sampling data, but only includes the relevant codes: "Lan" for landings above minimum reference size, "RegDis" for logbook registered discards and "BMS" for landings below minimum reference size.

Landings weight and value: In the data format the both the official and a scientific landings weight are included, and there is a field to indicate reason for the difference, if they are not equal.

- *Official weight*: The live weight in kg from official data.

---

[1] http://www.ices.dk/marine-data/data-portals/Pages/RDB-FishFrame.aspx

- *Scientific weight*: Live weight in kg which might include scientific estimates. This can be the same as the official weight.
- *Explain difference*: If there is a difference between the official weight and the scientific weight, it should be explained in this field. The field can be set to "Sample data" if sampling data are used to correct the official weights, "Unallocated catches" if there is knowledge about unallocated catches, "Area misreporting" if e.g. VMS data indicates fishing in another area than officially reported or "Correction for overweight in boxes" if sampling data indicates overweight in boxes.
- *Total landings value*: Sales value in Euro of the field official weight 'CLofficialWeight'. If nessesary the estimated value can be reported. Please report the data source in the field 'CLdataSourceOfScientificWeight'. If logbook registered discards, put NA..

Uncertainty indicators of estimated landings: Where estimates are given, there should also be an uncertainty indicator, which has been made possible in the data format:

- *Landings RSE*: The Relative Standard Error should be given if the weight of landings is estimated. If official data are used, it should be set to NA.
- *Value RSE*: The Relative Standard Error should be given if the value of the landings is estimated. If official data are used, it should be set to NA.
- *Qualitative bias*: If a quantitative RSE is not possible, a semi-quantitative scale ranging from +++ (large overestimate) to −−− (large underestimate) can be used as in Hyder et al 2017. If the weight of landings is based on official data, use NA.

Fields in Commercial Effort (CE) data format

The data format has been updated from the RDB to follow recommendations on effort calculations from the 2nd Workshop on Transversal Variables in Nicosia.

Data sources: In the dataset, it is possible to give both official and a scientific effort estimates. The scientific effort could be an estimate of the small-scale fishery based on sampling data, or it could be the official effort adjusted with e.g. VMS data. To make it transparent, it has been made possible to add information on data sources in the data format. Data should not be duplicated. The data sources are given in three fields:

- *Data type for scientific effort*: To be able to distinguish if data are based on official data (e.g. logbooks, sales notes, declarative forms) or it is estimates based on sampling (this could be the case for small-scale fisheries), the data type can be set to either "Official" or "Estimate".
- *Data source of scientific effort*: The source of the data can be set to "Logbook", "Sales notes", "Other declarative forms", "Combination of official data" or "Sampling data".
- *Sampling scheme behind scientific effort*: If sampling data are used to estimate the scientific effort, it should preferable be specified from the *Sampling scheme*. The national options of sampling from the *Sampling scheme* are maintained in a code list in the RDBES.

Number of trips: A trip is defined as the period between a vessel departs from a port (or factory ship) and arrives at a port (or factory ship). In the case of small-scale fisheries, one landing can equal one day at sea. The number of trips are reported as both:

- Number of fraction trips: If a trip covers more than one aggregation level of the primary key, it should be split up to the primary key according to effort, following the principles agreed on the 2nd Workshop on Transversal Variables on splitting up days at sea. This field makes it possible to split the effort in the areas where there was fished.

- Number of dominant trips: The trip is assigned to the dominant aggregation within the primary key (month, métier, ICES rectangle etc.) within the trip. This field will ensure one trip only contribute with 1 to the number of (dominant) trips.

The two fields are complementary. The sum of each field should be the same when aggregated over all areas and time, it is just a way to split the effort.

A simple illustrative example of these concepts is a 3-day fishing trip using the same active gear in two different statistical rectangles. The total number of days at sea for this trip is 3, with 2 of those days having fishing activity taking place:

Table 1 Fishing effort per day for a single trip

Day	Area	Rectangle	Metier	Fishing Time (hours)
1	27.7.g	32E2	OTB_DEF_70-99_0_0	5
2	27.7.g	32E2	OTB_DEF_70-99_0_0	2
2	27.7.g	31E3	OTB_DEF_70-99_0_0	3.5
3	No fishing activity			

We calculate the values for Days at Sea and Fishing Days for each fishing date according to the principles in Annex 5 of "Report on the 2nd workshop on transversal variables". In the following tables FD stands for Fishing Date, whilst GA stands for GearArea, in line with the transversal report.

First the Days at Sea are calculated for each fishing date (remembering the trip has a total of 3 days at sea):

Table 2 Days at sea calculation

Day	Area	Rectangle	Metier	FDprop	Number of GAs on that date	GAprop	GAFDprop	Days at Sea
1	27.7.g	32E2	OTB_DEF_70-99_0_0	0.5	1	1	0.5	1.5
2	27.7.g	32E2	OTB_DEF_70-99_0_0	0.5	2	0.5	0.25	0.75
2	27.7.g	31E3	OTB_DEF_70-99_0_0	0.5	2	0.5	0.25	0.75

Then the Fishing Days are calculated (remembering we only have active gear used on this trip):

Table 3 Fishing days calculation

Day	Area	Rectangle	Metier	Number of active GAs on that date	GAeffort (active)	GAeffort (passive)	Fishing Days
1	27.7.g	32E2	OTB_DEF_70-99_0_0	1	1	0	1
2	27.7.g	32E2	OTB_DEF_70-99_0_0	2	0.5	0	0.5
2	27.7.g	31E3	OTB_DEF_70-99_0_0	2	0.5	0	0.5

The number of Fractional Trips is calculated for each row as the proportion of the total Days at Sea it contributes (this is equal to the value of GAFDprop in our calculation of Days at Sea):

Table 4 Number of fractional trips calculation

Day	Area	Rectangle	Metier	GAFDprop	Days at Sea	Number of Fractional Trips
1	27.7.g	32E2	OTB_DEF_70-99_0_0	0.5	1.5	0.5
2	27.7.g	32E2	OTB_DEF_70-99_0_0	0.25	0.75	0.25
2	27.7.g	31E3	OTB_DEF_70-99_0_0	0.25	0.75	0.25

Note that the sum of the number of fractional trip values is equal to 1 since we are considering a single trip.

Now we can aggregate the fishing effort by area, rectangle, and métier and also calculate the number of dominant trips – the row with the most Days at Sea now has the value of 1 for the number of dominant trips. Only a single row within the trip can be allocated as the dominant trip – if you have multiple rows with the same value for days at sea then you should use a consistent method to break the tie e.g. use the tied row with the highest value for fishing days or fishing time as the dominant trip.

Table 5 Fishing effort for example trip aggregated by area, rectangle, and metier

Area	Statistical Rectangle	Metier	Fishing Time (hours)	Days at Sea	Fishing Days	Number of Fractional Trips	Number of Dominant Trips
27.7.g	32E2	OTB_DEF_70-99_0_0	7	2.25	1.5	0.75	1
27.7.g	31E3	OTB_DEF_70-99_0_0	3.5	0.75	0.5	0.25	0

Note that sum of the number of fractional trips equals the sum of the number of dominant trips (in this example 1).

For full details of the calculations for days at sea and fishing days please read Annex 5 from the "Report on the 2nd workshop on transversal variables, Nicosia, Cyprus, 22-26 February 2016 : a DCF ad-hoc workshop" https://op.europa.eu/en/publication-detail/-/publication/8c5583fa-c360-11e6-a6db-01aa75ed71a1

Effort measures: A number of effort measures: Days at Sea, Fishing Days, Number of hauls/sets, Vessel fishing hours, Soaking meter hours, kW Days at Sea, kW Fishing Days, kW Fishing Hours can be reported with both Official and Scientific values. They can be equal, but in some cases scientific

estimates can be reallocated (e.g. based on VMS data), or in the case of small scale fisheries, they can be based on other data sources.

The Soaking meter hours is used for passive gears as the total meters of net multiplied with the number of hours the gear is fishing. The field is optional, as it is not always known.

The Vessel fishing hours is used for active gears, but can also be used for the handling of passive gears. It is an optional field that can be based on logbook registrations, or VMS data.

The official data on Soaking meter hours, Vessel fishing hours and Number of hauls/sets might be only partially informed (especially for vessels under 10 meters that does not have logbooks with start and end time and is sometimes not filled in), but the scientific estimates should be for the full aggregation level. Information from VMS/AIS, sampling programmes and questionnaires might help informing the variables.

Number of unique vessels: Number of the active unique vessels within the aggregation level. The sum of unique vessels will be much higher than the actual number of unique vessels for the country, because one vessel can be counted many times if fishing in several months, statistical rectangles etc. This number should not aggregated, but can be used to know if data are confidential (individual vessels can be identified) or can published.

Uncertainty indicators of estimated effort: Where estimates are given, there should also be an *uncertainty indicator, which has been made possible in the data format:*

- *Effort RSE*: Relative Standard Error of the estimated effort. For official data: NA.
- *Qualitative Bias*: For estimated data, a semi-quantitative scale ranging from +++ (large overestimate) to −−− (large underestimate) can be used as in Hyder et al 2017.

Common fields for CE and CL data

This describes some fields that have been added in both the CL and CE data formats:

Exclusive Economic Zone Indicator (EEZI): Indicating if the landings are from EU waters (EU), International waters (RFMO) or non-EU coastal waters (COAST) this field can enable the data to be exported for the EU FDI data call in the future.

Incidental by-catch mitigation device: These can be coded as "Sorting grid", "Functional pingers", "Seal extruder device" and "Turtle extruder device", "Unknown" or "None". To support data for by catch.

Fishing Technique: This is an optional field to be used if the data are to be exported to the EU FDI datacall. It is the Fishing Technique codes defined for the EU AER Fleet Economic data call, and is also used by the EU FDI data call.

Appendix 3: Fishing technique coding

Fishing technique description	Code
Drift and/or fixed netters	DFN
Dredgers	DRB
Demersal trawlers and/or demersal seiners	DTS
Vessels using pots and/or traps	FPO
Vessels using hooks	HOK
Vessel using other active gears	MGO
Vessels using polyvalent active gears only	MGP
Vessels using passive gears only for vessels < 12m	PG
Vessels using other passive gears	PGO
Vessels using polyvalent passive gears only	PGP
Vessels using active and passive gears	PMP
Purse seiners	PS
Pelagic trawlers	TM
Beam trawlers	TBB
Inactive vessels (only for table J)	INACTIVE
No fishing technique (e.g. divers without fishing vessel)	NO

Dep Sea Regulation: This is an optional field to be used if the data should be export to the EU FDI data call. It is an indicator with options "Yes" or "No" if the fishery is under the deep sea regulation.

Description of each table

Upper hierarchy

Upper hierarchy table	Description	*be aware*
Design table (DE)	This table holds the information about the sampling scheme and stratification thereof. Further it specifies the upper hierarchy, which is unique for a given combination of sampling scheme and stratification.	
Temporal Event (TE)	If one of the sampling units is some kind of time e.g., minute, week or quarter, then this is where to put the information about the selection.	
Location (LO)	If one of the sampling units is a location, then this is where to put the information about the selection.	The location refers to a place with LOCODE.
Vessel Selection (VS)	If one of the sampling units is a vessel, then this is where to put the information about the selection.	This table is only used when the vessel is part of your sampling hierarchy. Information relating to the vessel properties can be found in Vessel Details (VD) table.
Fishing Trip (FT)	This table holds information about the fishing trip. Definition: Any voyage of a fishing vessel during which fishing activities are intended that starts at the moment when the fishing vessel leaves a port and ends on arrival in port.	Fishing activity does not necessarily imply catch. When observers are deployed on trips that e.g., do not register catch they should also be logged into the RDBES (e.g., gill net trips that only deploy the gear; purse-seine trips that search and do not find any fish)
Fishing Operation	This table holds information about the fishing operation – it is very similar to the HH in the	

(FO)	present RDB aggregated by haul. Definition: set/haul	
On-shore Event (OS)	If one of the sampling units is a combination of Location and Time, then this is where to put the information about the selection. Definition: Location*time - e.g. port-day, market-day, auction-day and factory-day.	
Landing Event (LE)	This table holds information about the landing event - it is very similar to the HH in the present RDB aggregated by trip, but it is not assumed that one is sampling from all the landings of a fishing trip because there may be e.g. landings and sales taking place elsewhere. Definition: Landings from a fishing trip or part thereof.	A fishing trip may register several landing events (e.g., land in two different ports). A landing event may be the product of several consecutive fishing trips made by a vessel. Landing event can also contain multiple gears and areas, in the table the dominant gear and area is use.
Species Selection (SS)	This table holds information about the selection of species expected in the Sample (SA) table e.g., only species from a national species list are sampled, full concurrent species sampling, probabilistic species selection from a species group.	This table is mandatory.
Sample (SA)	This table holds outcome of the sampling done in the upper hierarchy, i.e. the final sample of fish available for measurement or biological analyses. It is very similar to the SL in the present RDB.	This table is 'Self-referring': a single Sample (SA) table holds one level of subsampling, but it is possible to add more levels of subsampling by linking one Sample (SA) table to another with SAparentID.

Auxiliary tables

Upper hierarchy auxiliary table	Description	be aware
Sampling details (SD)	This table holds information on the sampling country and institute.	This table is mandatory.
Vessel details (VD)	This table holds information about the vessel properties e.g. homeport and vessel length	This table is mandatory. There are codes for unknown vessel if the vessel is truly unknown.
Species List Details (SL)	This table holds information about the list of species which are targeted during a sampling event.	This table is mandatory.

Lower hierarchy tables

Lower hierarchy table	Description	be aware
Frequency Measure (FM)	This table holds the length distribution of the subsample from Sample (SA) – very similar to the HL in the present RDB format.	This table holds the length distribution of the smallest subsample from the Sample table (SA).
Biological Variable (BV)	This table holds the biological measurement done on individual fish e.g., weight, age, maturity and stock. The measurements are very similar to the ones in the CS in the present RDB with added	The Single fish number (id) (BVfishId) is what links the different measurements on a single fish together.

	information about the selection, but the table has been transposed, so each line holds a single measurement. This facilitates adding new variables and allows for different protocols for selecting individual fish for different measurements e.g., age sampling may be length stratified, but weight sampling may not. Further it also makes it much easier to add new measurement types in the future.	Before the CA table in the present RDB was always linked to the TR table. The link is now specified by the upper and lower hierarchy.

Hierarchy Summary

The following diagram shows how the tables relate to the different levels of sampling units and the division between the Upper and Lower Hierarchies – the Vessel Details and Species List Details tables are omitted for clarity. As can be seen there is some duplication between the tables in the Primary Sampling Unit section and the Intermediate Sampling Unit section – this is because in some of the hierarchies a table might appear as a primary sampling unit whilst in others it might be an intermediate sampling unit.

Upper Hierarchy (can vary between sampling scheme and top level strata)	**Design Tables**	Design Table (DE) Sampling Details (SD)
	Primary Sampling Units (PSU)	Vessel Selection (VS)　　　Location (LO)　　　Fishing Trip (FT) Onshore Event (OS)　　　Temporal Event (TE)
	Intermediate Sampling Units	Vessel Selection (VS)　　　Location (LO)　　　Fishing Trip (FT) Onshore Event (OS)　　　Temporal Event (TE) Landing Event (TE)　　　Fishing Operation (FO) Species Selection (SS)
	Sample (and sub-samples)	Sample (SA)
Lower Hierarchy (can vary between samples)	**Data**	Frequency Measure (FM) Biological Variable (BV)

The following tables show a summary of the tables that appear in each of the upper and lower hierarchies. More details can be found in the worked examples in Annex I and in the hierarchy spreadsheet.

Table 1 Upper hierarchies

Only mandatory records are shown for the hierarchies below.

Hierarchy 1	Hierarchy 2	Hierarchy 3	Hierarchy 4	Hierarchy 5	Hierarchy 6	Hierarchy 7	Hierarchy 8	Hierarchy 9	Hierarchy 10
Design	Design	Design	Design	Design	Design	Design	Design	Design	Design
Sampling Details	Sampling Details	Sampling Details	Sampling Details	Sampling Details	Sampling Details	Sampling details	Sampling details	Sampling details	Sampling details
Vessel Selection	Fishing Trip	Temporal Event	On-shore	On-shore	On-shore	On-shore	Temporal Event	Landing location	Vessel Selection
Fishing Trip	Fishing Operation	Vessel Selection	Fishing Trip	Landing Event	Fishing Trip	Landing Event (b)	Vessel Selection	Temporal event	Temporal Event
Fishing Operation	Species Selection	Fishing Trip	Landing Event	Species Selection	Fishing Operation (a)	Species Selection	Landing Event	Species selection	Fishing Trip
Species Selection	Sample	Fishing Operation	Species Selection	Sample	Species Selection	Sample	Species selection	Landing Event (b)	Fishing Operation
Sample		Species Selection	Sample		Sample		Sample	Sample	Species Selection
		Sample							Sample

Hierarchy 11	Hierarchy 12	Hierarchy 13
Design	Design	Design
Sampling details	Sampling details	Sampling details
Location	Location	Fishing Operation
Temporal Event	Temporal Event	Species Selection
Fishing Trip	Landing Event	Sample
Landing Event	Species Selection	
Species Selection	Sample	
Sample		

(a) only FOaggregationLevel ==T

(b) LE is not a part of the estimation hierarchy, please see annex for the specific hierarchy 7.

Vessel Details and Species List

The Vessel Details have to be uploaded into a separate table before the different hierarchies can be uploaded. Then the Vessel Selection should just refer to the Vessel Details through the Encrypted Vessel Code. Other tables/records like e.g. Fishing Trip and Landing Event should also just refer to the Vessel Details

The Species List also have to be uploaded into a separate table before the different hierarchies can be uploaded. Then the Species Selection should just refer to the Species List name.

Upload order of the record types

Vessel Details, VD, and Species List, SL have to **be uploaded before the hierarchies** can be uploaded.

H[all]	H[all]
VD	SL

The following are the import orders of the record types in the RDBES data file by hierarchy. In the column header the 'H' stands for hierarchy, directly followed by the hierarchy number.

'-' (minus after the record type) means mandatory but not in the estimation hierarchy.

'()' (the record type in brackets) means optional and not in the estimation hierarchy.

H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13
DE	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE	DE
SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-	SD-
VS	FT	TE	OS	OS	OS	OS	TE	LO	VS	LO	LO	(FT)
FT	FO	VS	FT	(FT)	FT	LE-	VS	TE	TE	TE	TE	FO
FO	SS	FT	LE	LE	FO	SS	(FT)	SS	FT	FT	(FT)	SS
SS	SA	FO	SS	SS	(LE)	SA	LE	LE-	FO	LE-	LE	SA
SA		SS	SA	SA	SS		SS	SA	SS	SS	SS	
		SA			SA		SA		SA	SA	SA	

The hierarchies here are comparable with the design classes in WKPICS 2013

The WKPICS design classes A, B and C below does not refer to the A, B, C and D in the lower hierarchies in table 2

	Hierarchy 1	Hierarchy 2	Hierarchy 3	Hierarchy 4	Hierarchy 5	Hierarchy 6	Hierarchy 7	Hierarchy 8	Hierarchy 9	Hierarchy 10	Hierarchy 11	12	Hierarchy 13
Design class (WKPICS, 2013)	B	A	Not defined	C	C	C	C	Not defined	D	B	D	D	Not defined

Table 2 Lower hierarchies

	Lower Hierarchy A	Lower Hierarchy B	Lower Hierarchy C	Lower Hierarchy D
Tables in the lower hierarchy	Length frequency	Length frequency	Biological Variable	
	Biological Variable			

Typically each of an Institute's sampling designs will be mapped to a single Upper Hierarchy. For example an institute might determine that Upper Hierarchy 1 is the best match for its sampling-at-sea programme. Although the samples from this programme are all recorded using the same Upper Hierarchy they can actually use different Lower Hierarchies if required. This can be relevant if there are different ways of recording length frequency data and biological data within a sampling programme – for example, length data might be recorded at the haul level whilst biological measurements are recorded at the trip level. In this case the samples that are linked to individual hauls could use Lower Hierarchy B (which just stores length data) whilst those samples that are linked to the trip level could use Lower Hierarchy C (which just stores biological data). This could also be relevant if length frequency are taken on all species , but biological measurement are only taken on a specific group of species, these species will then have Lower Hierarchy A while the others will have Lower Hierarchy B.

The new design variables and hierarchies

Design variables (applies to all tables that are part of hierarchies)

The RDBES makes a strong push towards the specification of the core variables needed to describe the sampling and carry out design-based estimation. Those variables are termed Design Variables and are identified in the data model with the acronym "DV". Among the mandatory Design Variables (DV, M) are essential aspects of the design like stratification, clustering and selection methods. Optional Design Variables (DV, O) include aspects that, albeit essential for design-based estimation, are frequently not available in more ad-hoc designs such as the total number of units, the number of units sampled or the probability of inclusion.

Sampling hierarchies (applies to Design Table; Sample Table)

The designs used in the sampling of commercial fisheries in ICES waters are mostly, if not all, multi-stage. In multi-stage designs the final sample (e.g., the fish sampled) is selected through a set of stages where the sampling units at each stage are (sub-) sampled from the (larger) units chosen at the previous stage.

In the RDBES the **sampling hierarchy** is a set of instructions that indicates what sampling levels are included in the multi-stage sampling of the commercial catches and how they are (hierarchically) related to each other. In a sampling hierarchy each level of a multi-stage sampling design is represented by a hierarchy table that holds the units sampled and details on how that sampling was done. The RDBES is very flexible with regards to the sampling levels that can be included in a hierarchy, allowing for a wide variety of sampling units to be specified such as ports, vessels, trips, fishing operations, time, amongst other. The kind of sampling units supported are distilled from analyses of commercial sampling programs by a range of ICES expert groups and workshops. Less common sampling units are not supported with designated tables, but may be modelled through the use of columns that describe clustered sampling (see Section "The new design variables").

The sampling hierarchies are fundamental to the RDBES data model because they determine which information each country should upload to it. Two types of sampling hierarchies must be defined for each sampling programme. An **upper hierarchy** – defined in the "Design table" and that includes the upper levels of the design from "programme" to "fish sample". A **lower hierarchy** – defined in the "Sample Table" and that respects to the lower levels of the design, i.e., from "fish sample" downwards to length measurements and/or biological analyses carried out at the level of individuals.

Samples (applies to all tables that are part of hierarchies)

In the hierarchy tables of the RDBES a row represents a unit sampled at that level of the hierarchy. E.g., in the port table, each row represents a port sampled; in the vessel table each row represents a

vessel sampled; and so on. Consequently the number of rows in each hierarchy table generally corresponds to the number of elements in the sample taken from that level. An exception occurs when one or more strata were not sampled (see *stratification*).

Total units and Sampled units (applies to all tables that are part of hierarchies)

All tables involved in the hierarchies defined in the RDBES include columns for total units and sampled units - named numberTotal and numberSampled in all table with table name as prefix e.g. VSnumberTotal and VSnumberSampled in the Vessel Selection table. In general, the total column is defined as the total number of sampling units in that sampling stage; and the sampled column is defined as the number of units sampled in that sampling stage. In the case of stratification (see *stratification*) the totals and sampled respect to the total size and sample size of each stratum (see example 1 below). In the case of clustering (see *clustering*), total and sampled columns refer to the total and sampled number of units in each cluster *and* two new variables are used numberTotalClusters and numberSampledClusters (also prefixed with table name) that indicate the total (in the population) and sampled number of clusters. In the particular case of Selection method = "census" total and sampled are equal.

Stratification (applies to all tables that are part of hierarchies)

A column is included in every hierarchy table that allows the presence of stratification at each sampling level to be declared - named stratification in all table with table name as prefix e.g. VSstratification in the Vessel Selection table. When stratification exists, the strata names are declared in the stratum column (also prefixed with table name).

The vast majority of strata considered in the sampling of commercial fisheries have known size and are composed of countable units. Accordingly when stratification is implemented the tables of the RDBES require the specification of a) the total size of each stratum and b) the size of the sample is known. This is necessary because each unit sampled is only meaningful when related to the size of the stratum size and the sample it came from. Additionally it is important that every stratum is sampled (examples 1a and 1b).

In the RDBES stratum total and sample size values are expressed as a number of sampling units (e.g., number of vessels in the stratum and number of vessels sampled from that stratum). Depending on the table, other common units of strata size may also be available (e.g., weight).

Example 1a: Simplified example of unstratified vessel table. 7 samples were taken from the entire (non-stratified) population of vessels. Compare with 1b (stratified).

VSid	VSstratification	VSstratum	VSselectionMethod	VSnumberTotal	VSnumberSampled
1	N	U	random	1700	7
2	N	U	random	1700	7
3	N	U	random	1700	7
4	N	U	random	1700	7
5	N	U	random	1700	7
6	N	U	random	1700	7
7	N	U	random	1700	7

Example 1b: Simplified example of stratification by vessel size. In this case the aim of the sampling scheme was to estimate catches at fleet-level (i.e., for all vessel sizes). Note that for all strata, stratum totals and sample sizes are reported and that 3 samples were taken from stratum <10m and two from the remainder two strata (>=10 <15m and >=15m). Compare with 1a (no stratification).

VSid	VSstratification	VSstratum	VSselectionMethod	VSnumberTotal	VSnumberSampled
1	Y	<10m	random	1000	3
2	Y	<10m	random	1000	3
3	Y	<10m	random	1000	3

4	Y	>=10 <15m	random	500	2
5	Y	>=10 <15m	random	500	2
6	Y	>=15m	random	200	2
7	Y	>=15m	random	200	2

Guidelines for reporting stratification in RDBES:

- All strata included in the sampling frame of a particular sampling stage must be reported *even if they have not been sampled*. When a stratum has not been sampled it is reported as a line where selection method = "not-sampled" and sample size = 0.

   o E.g.:
      ▪ if your 1st sampling level is Vessel and the goal of your sampling scheme is to obtain data to estimate size composition for the entire catch of your country (i.e., all vessels) by using a vessel list stratified by vessel-size then all vessel size strata should be declared; In this case if size-class [>=15m] was not sampled it should be reported with selection method = "not-sampled" and sample = 0 (see example 2). The estimations for [>=15m] will be estimated from other samples

Example 2: Simplified example of stratification by vessel size with no samples taken in one stratum (>=15m). Compare with example 1 (stratified, all strata sampled) and 3.

VSid	VSstratification	VSstratum	VSselectionMethod	VSnumberTotal	VSnumberSampled
1	Y	<10m	random	1000	3
2	Y	<10m	random	1000	3
3	Y	<10m	random	1000	3
4	Y	>=10 <15m	random	500	2
5	Y	>=10 <15m	random	500	2
6	**Y**	**>=15m**	**not-sampled**	**200**	**0**

- Out of frame strata should not be declared.
   o E.g.:
      ▪ If the study population of the sampling programme is vessels >=10 m i.e., if you do not expand your sample to a total that includes vessels <10 m, there is no need to declare the strata <10 m (see example 3). This mean that there will not be any estimations for <10 m.

      ▪

Example 3: Simplified example of stratification by vessel size where one strata is out-of-frame (<10 m) and therefore does not need to be declared. Compare with example 2 where the VSid 6 with strata >=15 is in-frame but not sampled.

VSid	VSstratification	VSstratum	VSselectionMethod	VSnumberTotal	VSnumberSampled
1	Y	>=10 <15m	random	500	2
2	Y	>=10 <15m	random	500	2
3	Y	>=15m	random	200	2
4	Y	>=15m	random	200	2

- Note that in the sampling table, besides the stratum column, many other commonly observed variables are defined some of which are sometimes used to stratify fractions of the

catch (e.g., size category); However, following a similar procedure to all other tables, it is important that the full stratification is declared in the stratum column as this will be the one considered for effects of partitioning the population (see Example 4). Also, in agreement with the general stratification guidelines (see above) all size categories present (e.g., in a landing) should be reported in the sample table even if they were not sampled *but* size categories absent from that landing need not be reported (see Example 4).

Example 4: Simplified example of stratification in the sample table. Note that stratum indicates stratification was based on Commercial size category but not on presentation; and that only 4 size categories were actually present in the landing. Other Commercial size category (e.g., 5, 6) were absent from the landings.

SAid	SAcatch-Fraction	SAspecies-Code	SAstratum	SAcomm-SizeCat	SApresen-tation	SAnumberTotal	SAnumberSampled
1	Lan	126436	Size 1	Size 1	Whole	1	1
2	Lan	126436	Size 2	Size 2	Gutted	5	2
3	Lan	126436	Size 3	Size 3	Gutted	10	2
4	Lan	126436	Size 4	Size 4	Gutted	30	2

Clustering: One stage and Two stage cluster sampling (applies to all tables that are part of upper hierarchies, except sample)

The RDBES can accommodate several types of one- and two-stage cluster sampling of, e.g., clusters of ports within geographical regions, clusters of weekdays within months or quarters, vessels, etc. The specification of cluster sampling in the RDBES follows the same line of reasoning as the specification of stratified sampling and both can be combined to accurately represent the way sampling units are actually selected. The difference is that when clustering is declared an additional total and sample value has to be specified that represents the total and sampled number of clusters (e.g. LOnumberTotalClusters, LOnumberSampledClusters). Several detailed examples of different types of 1-stage and 2-stage cluster sampling can be found in Annex VI. Note: the cluster sampling of units present in some tables of the RDBES is similar in nature to the cluster sampling already represented in the multi-stage nature of the upper hierarchies. Users are directed to explore first the possibility of representing the clustered nature of their design by the hierarchies already available; and only make use of the within-table cluster option in the cases where the clustered nature of their design cannot be represented otherwise (e.g., a succession of temporal event tables (TE) would be needed, such as when months are randomly selected and then weeks selected within months; a succession of location tables (LO) would be needed, such as when regions are randomly selected and then markets selected within regions). See more details in annex VI.

Selection and inclusion probabilities (applies to all tables that are part of hierarchies)

Each table of an RDBES hierarchy contains four probability columns, SelectionProb, InclusionProb, SelectionProbCluster and InclusionProbCluster. These columns should hold the specific selection and inclusion probabilities of each row/sampled unit. These probabilities are fundamental for design base estimations and not always can they be estimated from number sampled (numberSampled - n) and total number (numberTotal - N), see annex V.

Sample selection methods (applies to all tables that are part of hierarchies)

The RDBES allows data-submitters to specify the selection method they used when selecting samples in selectionMethod. The following methods can be specified:

i) Probability selection methods: the selection of the units is based on the theory of probability -at any stage of the operation of selection the probability of any set of units being selected is known.

- **Simple Random Sampling With Replacement (SRSWR):** random sampling of equally probable elements with replacement.
- **Simple Random Sampling Without Replacement (SRSWOR):** Simple random sampling of equally probable elements without replacement.
- **Unequal Probability Sampling With Replacement (UPSWR):** random sampling of unequally probable elements with replacement.
- **Unequal Probability Sampling Without Replacement (UPSWOR):** random sampling of unequally probable elements without replacement.
- **Systematic Sampling (SYSS):** Sample is obtained by a systematic method (as opposed to random choice) with a random starting point. E.g., sampling from a list by taking units that are at equally spaced intervals.
- **Census (CENSUS):** the sample corresponds to the full set of units in the population

ii) Non-probability selection methods: the selection of units is based in factors other than random chance, e.g., convenience, prior experience, voluntary registration, or the judgement of a researcher.

- **Quota Sampling (NPQS):** the selection of units is not probabilistic and it takes place until a pre-assigned number of units is attained (e.g., 200 otoliths of a species in a year, 3 trips samples on-board per quarter).
- **Expert judgement (NPEJ):** The selection of units is not probabilistic and takes place based on the individual judgement according to a set of clearly defined and documented rules (e.g., existing knowledge of a fishery)
- **Ad hoc sampling (NPAH):** The selection of units is not probabilistic and takes place without consistent guidelines.

iiI) Other

- **Not Applicable (NotApp):** Used only when filling in the selection method of auxiliary tables (e.g., Fishing Trip table under Hierarchy 5)
- **Not sampled (NotSam):** Used in cases where specific strata are not sampled.

Sample Weights

The hierarchy tables of the RDBES do not incorporate sampling weights as these can be calculated from other available data (e.g., numberTotal and numberSampled if random sampling; inclusion probabilities in the case of sampling with unequal probability).

Catch category

The Catch category is categorising the catch. But since the e.g. sea sampling is done before a catch is landed, not all the catch categories are the same for sampling data and landing data. The common categories are:

Catch categories for both sampling data and landing data:

- o 'Lan' – Landing
- o 'BMS' – Below Minimum reference Size

Sample (SA) additional Catch categories:

- o 'Dis' – Discard
- o 'Catch' – Catch (landing, BMS and discard not split)

Landing (CL) additional Catch categories:

- o 'RegDis' – Logbook Registered Discard

The following is an explanation of the catch categories:

Landing, 'Lan'

The 'Landing' Catch category will cover the scientific estimates of landing, which are categorised as above minimum size. Landings does not include BMS.

BMS Landing, 'BMS'

Relevant to stocks under landing obligations. The BMS landings consist of fish and crustaceans Below Minimum Size, as registered in the logbook or as estimated by fishery observers.

If it is possible to separate BMS and discards fractions when sampling from e.g. at sea observer programme, then the BMS estimate should be included. If it's not possible to separate discard and BMS fractions, then everything will be seen as discard.

Discard, 'Dis'

The 'Discard' Catch category will cover the discard fraction based on fishery observer estimations. This category is the part of the catch, which is thrown overboard into the sea.

Data for this fraction should be reported even when discard values are low. Discard estimations for pelagic species based on demersal observer programs should also be reported. This is especially important for some small pelagic stocks.

Catch, 'Catch'

When landing, BMS and discard are not split, it is possible to use the Catch category 'Catch' for the combined catch.

Logbook Registered Discard, 'RegDis'

This component corresponds to discards which are registered in the logbook.

Catch categories – under Sample SA

Sampling Catch categories are shown below. A sample of the catch can potentially be split in the following Catch categories; Landing, BMS and Discard should be sum to the total catch and should therefore not be overlapping.

If a catch does not contain a BMS landing part, then the BMS catch category should be ignored.

Sampling of catch can be split in the following Catch categories.

Sampling of a catch can be split in

| Landing

Landed catch categorised as above minimum size | BMS landing Landed catch categorised as below minimum size, BMS. | Discard

From sea observers | OR | Catch

Landing, discard and BMS is or cannot be split |
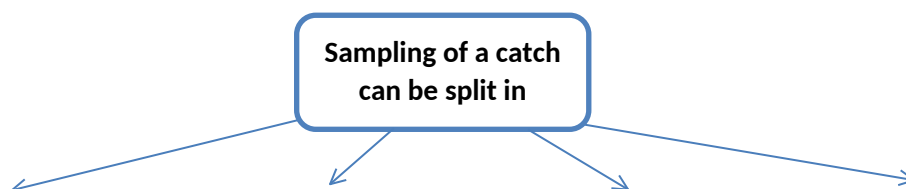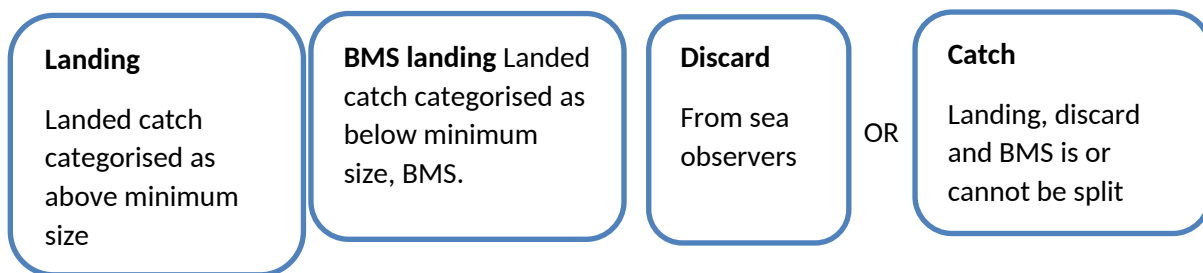
Catch categories – under Commercial Landing CL

Landing Catch categories are shown below. A landing can be split in the following Catch categories, which should sum to the total registered catch and should there for not be overlapping.

A registered catch can be split in the following Catch categories.

A registered catch can be split in

| Landing

Landed catch categorised as above minimum size | BMS landing Landed catch categorised as below minimum size, BMS | Logbook Registered Discard (including de minimis) and damaged fish |
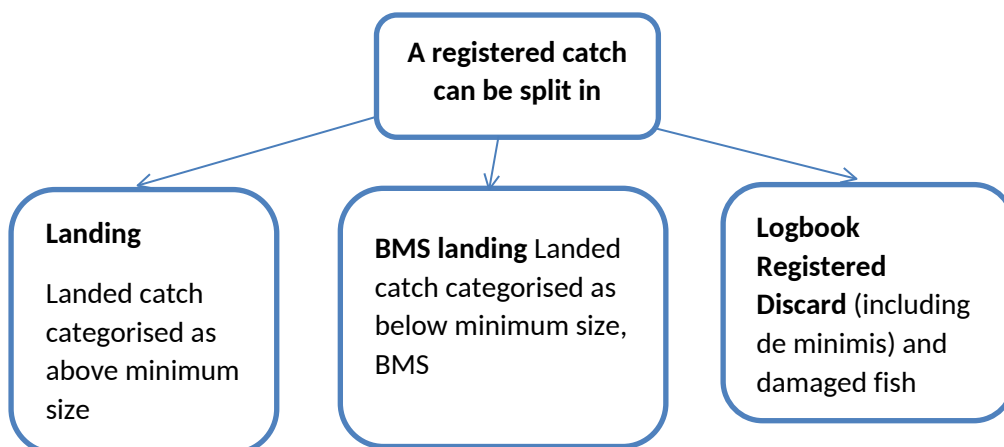
New features of the RDBES

True zero values and Non-recording zero values in species sampling

Not every sampling programme always records the weights or numbers of all species found in catches: cases exist where records are only kept a subset/list of taxa and no information is recorded from the remainder; in some extreme situations a sampling plan may record data from a single species despite the multi-species nature of the fishery. Additionally, sampling programmes may not routinely record specific groups (e.g. PETS - Protected Endangered Threatened Species). Such diversity of sampling protocols greatly complicates the estimation and may lead to very significant biases in, e.g., discard rates or biodiversity measures, when data analysts cannot readily distinguish true absences of a species from its routine (or sporadic) non-recording in database records.

The RDBES provides for the accurate distinction between true absences of a species and its non-recordings by means of a Species Selection table. This table sits directly above the sample table. In it, the species sampled *and* the list of species intended to be sampled are declared for each sampling

event (e.g., a landing or a fishing operation of a fishing trip). Some generic species lists are available (e.g., "all species including invertebrates") but sampling programme specific species lists can also be specified (e.g., a list of DCF species; a list containing only cod; a port specific list).

Non-responses and missing values due to quota sampling

The present version of the RDBES data model is prepared to receive non-responses resulting from a variety of causes, e.g., industry refusals, observer declines, etc. Full report of these occurrences is important for estimation because in many cases they cannot be assumed to be missing at random. The main tables of the present version of the RDBES contain a noSampReason that allows users to specify the main motives for not having sampled specific units - also prefixed with table name. The aim of this column in, e.g., the Fishing Trip Table, is to allow the calculation of refusal rates; In the Sample table, it will put in evidence, e.g., when biological samples were not taken due to quotas having been reached in, e.g., previous hauls - reaching of quarterly quotas (in the Sample table).

Example 1: Simplified example of reporting a refusal. In this case a sample of n=5 units was drawn (e.g., vessels contacted under hierarchy 1) but data could not be obtained for one of them due to a refusal. As a consequence VSid=5 will have no child entry.

VSid	VSselectionMethod	VSnumberTotal	VSnumberSampled	VSnoSampReason
1	random	10	5	
2	random	10	5	
3	random	10	5	
4	random	10	5	
5	random	10	5	Non-response – Industrial decline

Example: Simplified example of reporting of a non-sampling event due to quota being reached. In this case the cod was not biologically sampled because a quota had been reached in previous hauls. As a consequence SAid=1 will have no child entry in table Biological Variable, while SAid 2 and 3 will have child entries.

SAid	SAspeciesCode	SAnumberTotal	SAnumberSampled	SAnoSamp-ReasonFM	SAreasonNot SampledBV
1	Cod	10	2		Non-response – Quota reached
2	Hake	10	5	Non-response – sampling Error	
3	Haddock	10	5		

Recording of incidental by-catches and slipping

The present version of the RDBES is prepared to receive the information needed to estimate incidental by-catches and record drop-outs and slipping. With regards to incidental by-catches, the aim was to a) be able to accurately record positive incidental by-catch events (e.g., by extending the range of available taxonomic codes to marine mammals and birds), b) facilitate the correct distinction between non-observations (= missing values) and zero-observation (true 0s), c) accommodate the usage of different sampling methodologies (e.g., in by-catch studies it is frequent for observers to do a visual screening of the catch; in discard studies it is more typical for observers to collect weight-based samples) and d) be able to record the state of individuals (e.g., including codes for dead, wounded, alive, unknown, etc.). With regards to slipping and drop-outs the aim was to facilitate its reporting in a way that is clearly distinguishable from other fractions such as landings and discards. To accommodate these improvements the following changes were implemented:

- FO Table:

- o Inclusion of variable "Observation code" to indicate which part(s) of the fishing operation were effectively screened during the fishing operation. This new variable accomodates "Slipping" (Sl), "Drop-outs" (Dr), "Presorting" (Pr), "Sorting" (So) and combinations thereof.
- SS Table
  - o Inclusion of a variable "Observation_activity" (text), defined as SLIP (= SLIPPING), DROP (= DROP-OUTS), HAUL (= HAULING) and SORT (="SORTING"). This new variable allows the specification of different species lists for the different parts of the fishing operation, i.e., it is possible for some species to be screened for only at cod-end or different taxonomic levels to be used in observations of slipping and sorting of the catches.
  - o Inclusion of a variable "Observation type" (integer), defined as vol (= volume) and vis (= visual). This new variable allows for the consideration of different species lists in volume samples from those used in visual screening samples, ensuring the separation of the estimates obtained by the two different methods.
- SA Table
  - o Inclusion of a variable "SpecimenState" with a code list specifying the state of the individuals in the sample (e.g., dead, damaged).

Examples of how these new variables can be specified are present in annex IV of this report.

Other key concepts

Commercial species

A frequent need of commercial sampling programmes is that the species composition of boxes from generic commercial designations[2] (such as "rays", "anglerfishes", "seabreams", etc.) can be recorded. The RDBES is prepared to accommodate those data by offering the possibility of i) stating lists of commercial species in the Species Selection Table (SLid in the Species Selection Table) and ii) declaring, e.g., the sampled boxes of commercial species in the Sample Table (using the Commercial Species (SAcommercialSpecies) to identify the commercial name) with the biological species sampled featuring as subsamples.

Example: Simplified example of reporting of commercial species in table Species List Details (CHECK).

SLid	SLspeciesCode	SLcommercialSpecies
1	126554	MNZ
2	126555	MNZ

Recording of samples collected by fishers or control (new in the RBDES)

Data collection by fishers (Self sampling) or control agencies can take place at different levels of a sampling design: e.g., vessels are in some cases selected by the industry; the discards can be collected at-sea by fishers or control agents and processed later in the lab; in some cases, fishers themselves collect length frequencies and/or ages. Identification of cases when direct observation or validation by scientific observers did not take place is important so by default the main tables of the RDBES include a sampler variable that can be used to highlight such cases - named sampler in all table with table name as prefix e.g. VSsampler in the Vessel Selection table

---

[2] hereby termed commercial species for sake of simplicity

Recording of non-probabilistic sampling (new in the RBDES)

The RDBES is structured to accommodate statistically sound sampling designs. Such designs are generally composed of clearly delineated sampling hierarchies and probabilistic sample selection methods. Practice shows that is rarely the case but that in most cases, users are still able to relate their sampling protocol to one of the hierarchies and can record their departures from statistical principles in the selection method variable (e.g., as ad-hoc). However, in some cases (particularly in the case of poorly documented older data) it may be difficult to unequivocally identify the levels of the design themselves, i.e., different hierarchies may appear to equally fit the data. In such cases it is recommended that countries make a best guess and use the hierarchy they think most approximate the original one. Then, variable DEhierarchyCorrect in the design table should be used to identify that those hierarchies are ambiguous and likely mis-specified.

Annex I: Description of each hierarchy

The following is a short description of each of the 13 upper hierarchy and the 4 lower hierarchies. The 1, 2, 3, 10 and 13 of the 13 upper hierarchies are often used for at-sea sampling, the others are often used for on-shore sampling. Remember before uploading any hierarchies the Vessel Details and the Species Lists have to be uploaded, so the tables/records in the hierarchies can refer to the Vessel Details and the Species Lists.

Hierarchy 1 Sampling from a vessel list or from a reference fleet

Hierarchy 1 is most used for at-sea sampling, but can be used for on-shore sampling as well.

Different schemes which all fits the hierarchy 1:

- at-sea sampling with vessel selection from a vessel list, then fishing trip(s) from each vessel. Some or all fishing operations (hauls/sets) are sampled;
- at-sea sampling with vessel selection from a reference fleet, then fishing trip(s) from each vessel. Some or all fishing operations (hauls/sets) are sampled
- at-sea sampling with vessel selection from a vessel list, then fishing trip(s) from each vessel. Some or all fishing operations (hauls/sets) are sampled; some samples come from individual fishing operations (aggregation H) and some from the whole trip (aggregation T)
- on-shore sampling with vessel selection from a vessel list, then fishing trip(s) from each vessel. The last fishing operation (haul/set) is sampled.

Upper hierarchy tables:

- Design (DE)
- Vessel Selection (VS)
- Fishing Trip (FT)
- Fishing Operation (FO)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory) When imported VD has the link from FT, and is therefore after VS and before FT
- Species List Details (SL) (mandatory)

Legend
- Mandatory table in hierarchy
- Mandatory table not in hierarchy
- Auxiliary table not in hierarchy
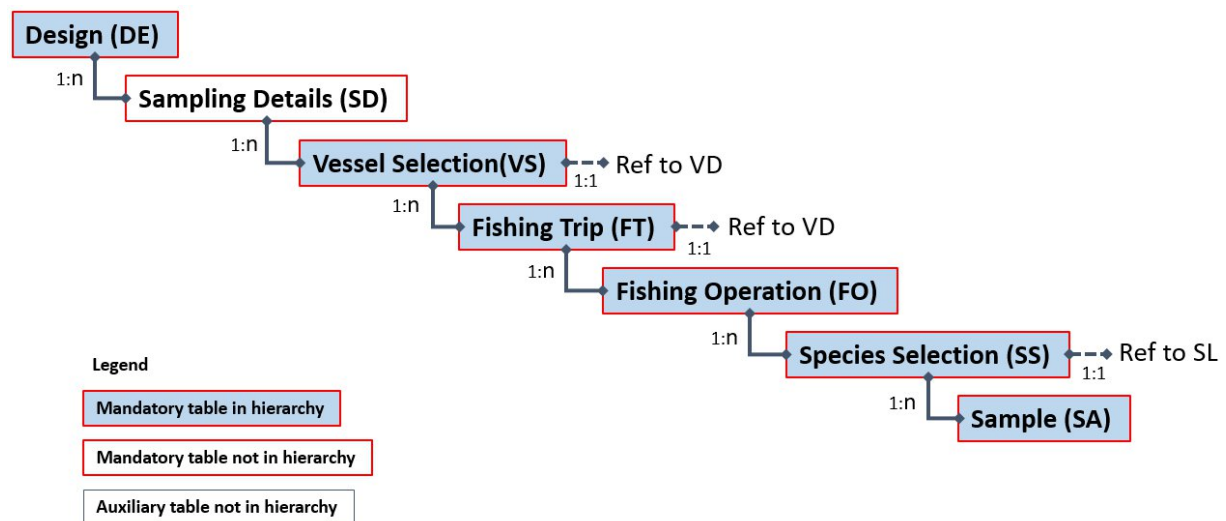
Example(s) described:

For example, an institute has set up a reference fleet for gillnetters in a specific area. The 1st of January 5 vessels are selected. During the year a number of fishing trips are sampled on each vessel. The selection of fishing trips is stratified by quarter. Data from one fishing trip is shown in the example. During the fishing trip one fishing operation is carried out. All species are weighted in the catch (no species selection). Catch is stratified by catch category.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Another example, an institute has set up a reference fleet for gillnetters. At the beginning of each quarter a list of vessels is being selected from the gillnetters reference fleet. Afterwards a contact person for each vessel is being called and asked about the possibility of on-shore sampling. In case of agreement, a fisherman gives the details about planned harbour and time of arrival. An observer is being sent to the given place on a given day and takes the sample from the last haul. All species are weighted in the catch.

Hierarchy 2: Sampling from a list of trips

Hierarchy 2 is most used for at-sea sampling, but can be used for on-shore sampling as well.

Different schemes which all fits the hierarchy 2:

- at-sea sampling from available fishing trips; some or all fishing operations (hauls/sets) are sampled
- at-sea sampling from available fishing trips; aggregated fishing operations (hauls/sets) are sampled
- on-shore sampling from available fishing trips; aggregated fishing operations (hauls/sets) are sampled when landed
- on-shore sampling from available fishing trips; the last fishing operation (haul/set) is sampled.

A list of trips to sample from is not always available, but a vessel list is used as a proxy for selecting trips. In reality we often try to squeeze sampling design where we sample from a vessel list into this hierarchy, when it may be more appropriated to use hierarchy 1[3].

Upper hierarchy tables:

- Design (DE)
- Fishing Trip (FT)
- Fishing Operation (FO)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)
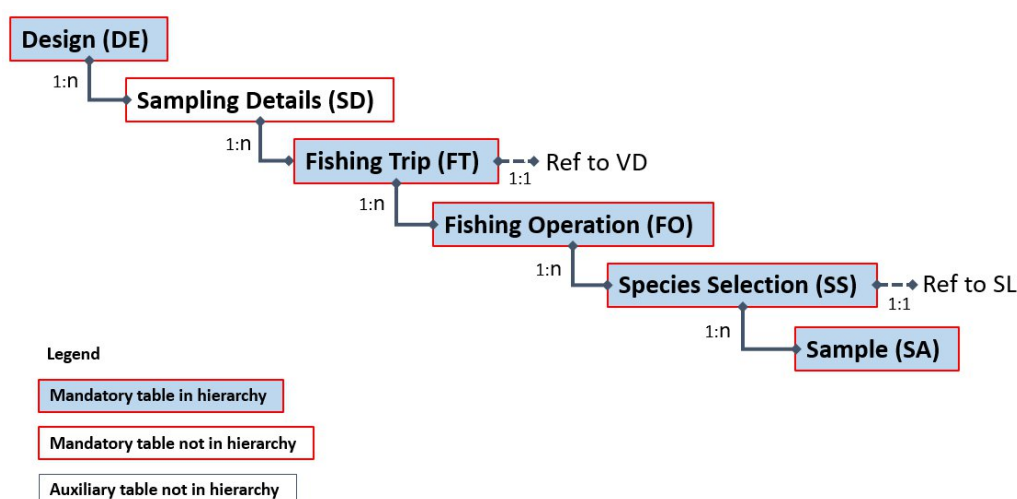
Example(s) described:

---

[3] "**At-sea sampling with trips as primary sampling units**. When trips can be selected randomly from a fleet of vessels, at least approximately, it is often reasonable to treat vessel-trips as the primary sampling units. In such cases, the list of all trips (obtained at the end of the year) makes up the sampling frame. This is a virtual frame that cannot be used in stage 1 to select the trips. The actual selection is typically based on a frame with a vessel list crossed with time. For a fleet with day-trips this can easily be achieved by randomizing the selection of days and vessels. For fleets with varying trip-length it is more difficult to selected vessels and trips with approximately equal inclusion probabilities. It can be helpful to create strata where vessels with a similar trip length are grouped", (WKPICS, 2013 p. 41)

For example, an institute samples a fishing trip conducted by a vessel from the "Hirtshals-TBB-NA-1" stratum. During the fishing trip 26 fishing operations are carried out (one of them being invalid and therefore not included in 'FOnumberTotal') – 6 of the valid hauls are sampled without stratification. All species are weighted in the catch (no species selection). Catch is stratified by catch category and species are sampled for weight on these hauls.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 3: Sampling where time (e.g., days, weeks) is the primary sampling unit and vessel is the second sampling unit

Hierarchy 3 is most used for at-sea sampling, but can be used for on-shore sampling as well.

Different schemes fit the hierarchy 3:

- At-sea sampling where a time period is first selected (e.g., a set of weeks) for sampling to take place. Then, in each time period, vessels in a list are contacted and fishing trips booked. Observers are deployed on fishing trips and sample the fishing operations. In each fishing operation the two catch categories - discards and landings – that are sampled separately and the weights of the species present in each category are determined.
- Variants:
    a. Self-sampling by fishers
    b. Selection of different time units (e.g., day, month)
    c. Sampling all fishing operations or a subset of them
    d. Sampling only discards, taking a catch sample that is sorted later into landings and discards, etc.
    e. Sampling all species or only species from pre-defined list(s)
    f. Different sampling and subsampling strategies for landings and discards (e.g., stratification by size category, subsampling by sex, etc.)

Upper hierarchy tables:

- Design (DE)
- Temporal Event (TE)
- Vessel Selection (VS)
- Fishing Trip (FT)
- Fishing Operation (FO)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory) When imported VD has the link from FT, and is therefore after VS and before FT
- Species List Details (SL) (mandatory)

Legend
- Mandatory table in hierarchy
- Mandatory table not in hierarchy
- Auxiliary table not in hierarchy
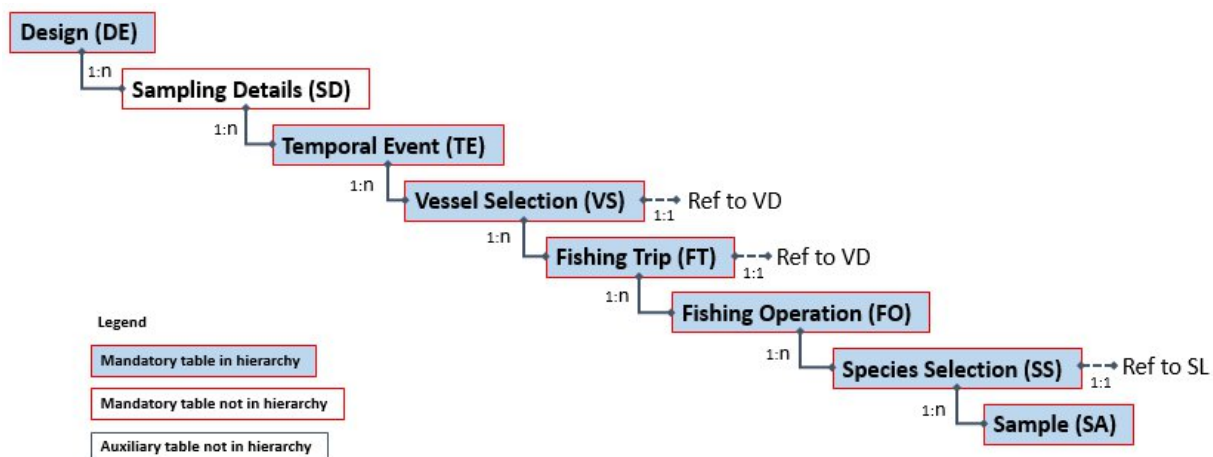
Example(s) described:

In the sampling design of "A Country" (XYZ), an institute (ANI) selected 3 weeks randomly each quarter for at-sea sampling to take place. In each selected week, 2 vessels were randomly selected and contacted. In a specific week all vessels contacted agreed to cooperate. For each vessel a trip was randomly selected from the ones the vessel planned to undertake that week and observers deployed on it. Observers sampled all fishing operations (n=6) and sampled both the landings and the discards of each fishing operation. The landings were sorted by the fishers into species*size category and the observers used boxes to subsample them. The discards were originally unsorted in a bulk of 13 containers and observers randomly sampled one of them. The content of that container was sorted by species and each species was subsampled to meet pre-specified goals (quota sampling).

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 4 Sampling by selecting from location*time (e.g. harbour-day), then from Fishing Trips, then from Landing Events from those Fishing Trips

Hierarchy 4 is most used for on-shore sampling, but can be used for at-sea sampling as well.

Different schemes which all fits the hierarchy:

- On-shore sampling at markets or ports. A location and time period (e.g. harbour-day) are chosen and then Fishing Trips that are landing at that location-time period are sampled; detailed information on the fishing trip is available, then Landing Events from those sampled Fishing Trips are sampled.

Upper hierarchy tables:

- Design (DE)
- On-shore Event (OS)
- Fishing Trip (FT)
- Landing Event (LE)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)
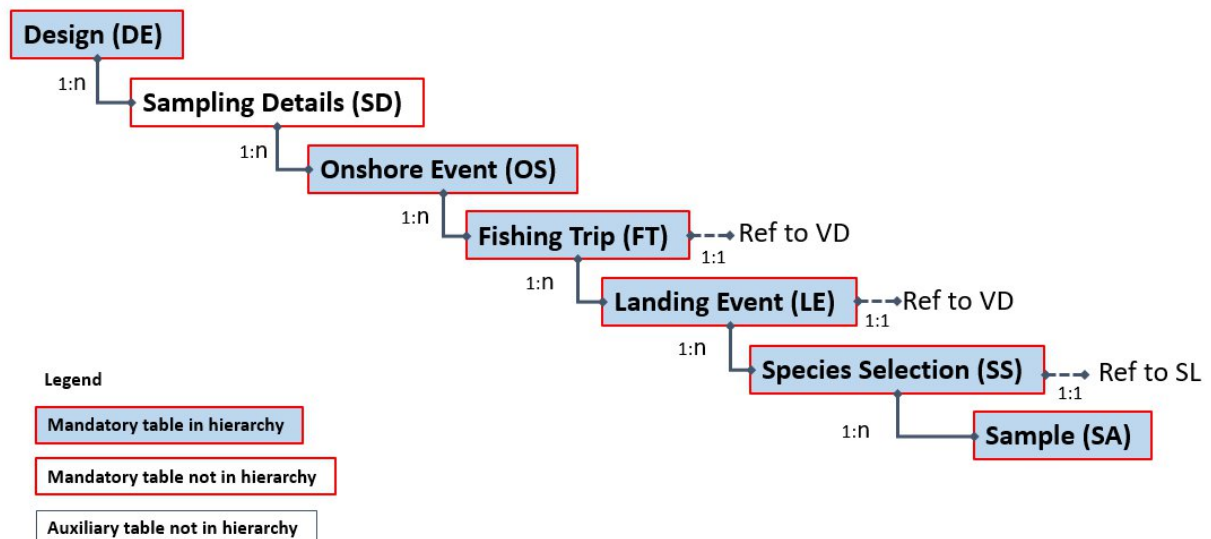
Example(s) described:

An institute has a port-week as its primary sampling unit. For example, an institute samples Week 3 at "Pelagic Port A" in the "PelagicQ1" stratum. During that week 5 fishing trips return to port (it is known that these fishing trips will land their entire contents at this port during this week) – these trips are unstratified and 2 of them are sampled. 1 of those fishing trips (PTrip1) has a single landing

event, but 1 of them (PTrip2) lands over 2 days. The single landing event of PTrip1 is sampled, but only 1 of the 2 landing events for PTrip2 is sampled.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 5 Sampling by selecting from location*time (e.g. harbour-day) as primary sampling unit, then from Landing Events as the secondary sampling unit.

Hierarchy 5 is most used for on-shore sampling, but can be used for at-sea sampling as well.

Different schemes which all fits the hierarchy 5:

- On-shore sampling from a list of landing sites sampled at random days. In each day observers sample some of the landing events present on-site. Data on the associated fishing trips is generally known, but not part of the selection procedure.
- On-shore sampling from a list of landing sites sampled at random days, samples of all or some landings. Fishing trip is not known.
- Variants:
    a. Random selection from a list of landing sites.
    b. Random selection from a cluster of landing sites, followed by random selection within clusters.
    c. Sampling surveys visiting landing sites along coast, with randomized start times and start positions.

Upper hierarchy tables:

- Design (DE)
- On-shore (OS)
- Landing Event (LE)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD)
- Vessel Details (VD) (mandatory)
- Fishing Trip (FT) (optional)
- Species List Details (SL) (mandatory)

Legend:
- Mandatory table in hierarchy
- Mandatory table not in hierarchy
- Auxiliary table not in hierarchy
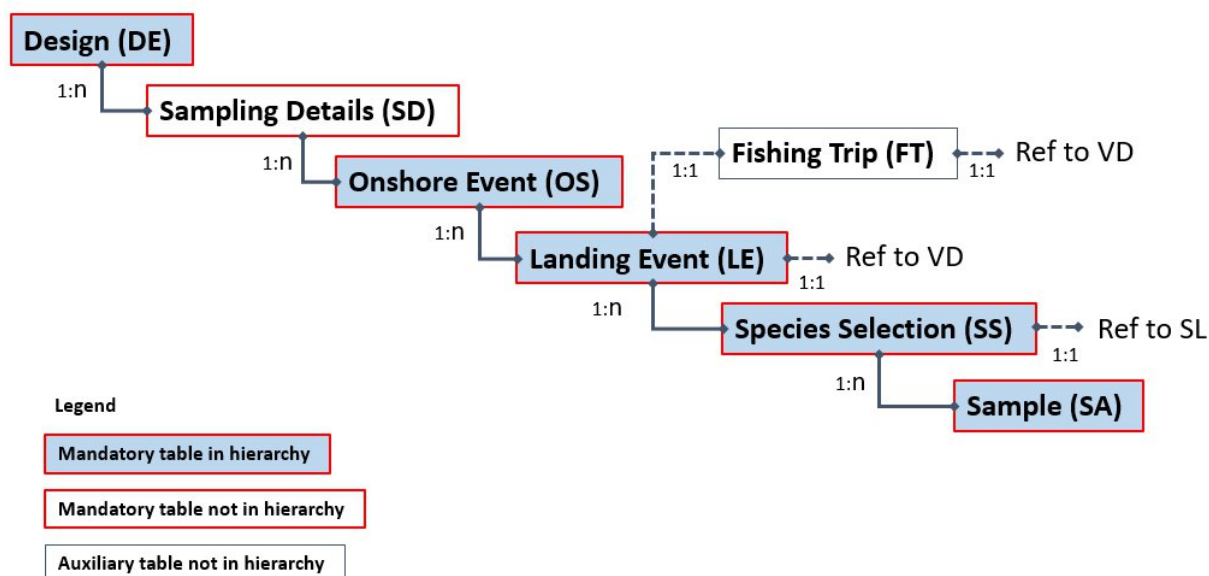
Example(s) described:

Sampling days are selected randomly, and on each selected day, a landing site is selected at random from a list of landing sites. Each site-day is registered in the *Onshore Event* table and visited by observers. In each visit, there are many landing events on site and some of them are selected for sampling and recorded in the *Landing Event* table (Note: the observers just know that some landings from a trip are there, they cannot guarantee that each landing event effectively comprises the entire landings from that trip, or that it does not accumulate fish from two successive trips from a single vessel). In each landing event the species present in a particular list of target species are sampled (the list is specified in the Species Selection table). For each target species the corresponding fraction of the landing is registered as a sample in the *Sample* table, and then a subsample of 30 fish registered in the same table.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 6 Sampling where Fishing Trips are sampled

Hierarchy 6 is most used for on-shore sampling, but can be used for at-sea sampling as well.

Different schemes which all fits the hierarchy 6:

- On-shore sampling at markets or landing ports; known Fishing Trips sampled at random; sale categories and single, concurrent or stratified species selection.
- On-shore sampling from buyers or vendors; known Fishing Trips; sale categories of aggregated landing, and single, concurrent or stratified species selection.
- Note: Hierarchy 6 differs from Hierarchy 7 on the certainty of information existing on the trips and samples. The trips of hierarchy 6 generally involve a single gear and fishing area and their landing events are generally known to represent the entirety of catch (i.e., no catch retained on board these trips). Hierarchy 7 is much more common and should be used where there is the possibility that on-shore samples come from different gears or areas and/or where there is the possibility that samples obtained do not represent the entire trip (e.g., if some retained catch can be kept on board for future landings; transhipment may have taken place, etc.). See FAQs.

Upper hierarchy tables:

- Design (DE)
- Sampling Details (SD)
- On-shore Event (OS)
- Fishing Trip (FT)
- Fishing Operation (FO)*
- Species Selection (SS)
- Sample (SA)

* FOaggregationLevel must be set to "T"

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Landing Event (LE) (optional)
- Species List Details (SL) (mandatory)
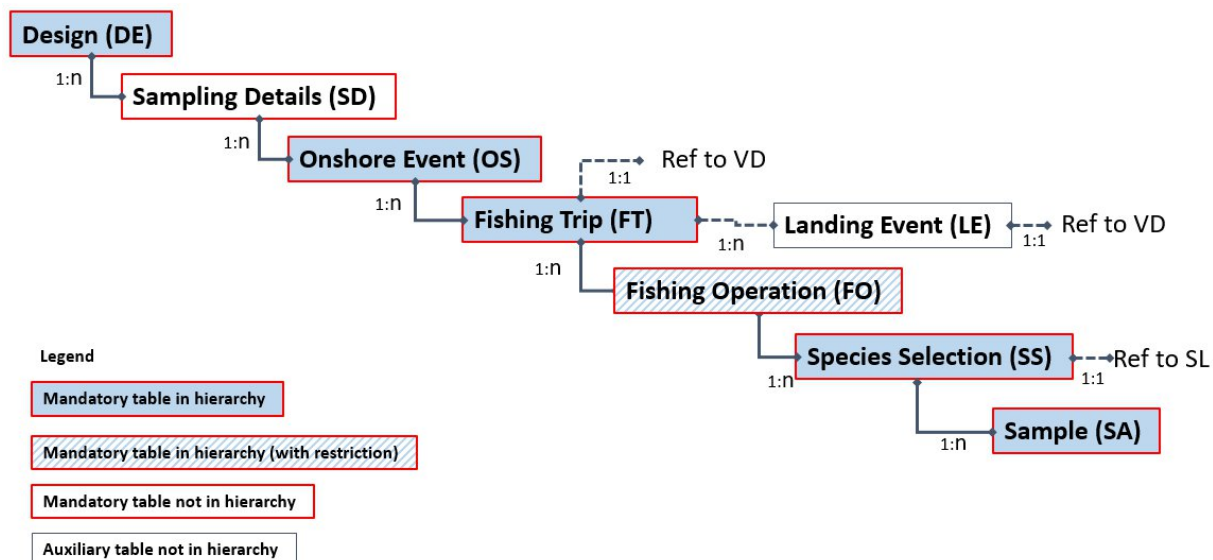
Example(s) described:

an institute has a port-day as its primary sampling unit. For example, an institute samples Day 28 at "Port DFG" in the "SmallPortQ1" stratum. During that day 6 fishing trips return to port (it is known that these fishing trips fished in a single area and with a single gear and that they land their entire contents at this port during this day). 2 of these trips are in the "DayTrip" stratum, and 4 are in the "LongerTrip" stratum. All landing events from 1 of the trips in the "DayTrip" stratum are sampled, and all landing events from 1 of the fishing trips in the "LongerTrip" stratum are sampled.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 7. Sampling at markets or ports, where location*time is the primary sampling unit

Hierarchy 7 is most used for on-shore sampling, but can be used for at-sea sampling as well.

On-shore sampling at markets or ports, where location*time is the primary sampling unit and species are the secondary sampling unit. Fishing Trips and landing events may be known but are not a part of the selection.

Different schemes which all fits the hierarchy 7:

- On-shore sampling at markets or ports. Some or all species present at the market are sampled without considering which fishing trip or landing event they come from. Information on the fishing trip and/or landings events may be collected but is auxiliary.
- On-shore sampling at markets or ports. Information on the fishing trip or landings event is not available. Some or all species of the landing are sampled.
- Variants:
  a. On-shore sampling from buyers or vendors.
  b. Sampling boxes from size categories

Upper hierarchy tables:

- Design (DE)
- On-shore Event (OS)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Landing Event (LE) (mandatory)
- Species List Details (SL) (mandatory)

Design (DE)
1:n
Sampling Details (SD)
1:n
Onshore Event (OS)
1:n
Species Selection (SS) ← --→ Ref to SL
1:1
1:1
Landing Event (LE) ← --→ Ref to VD
1:1
1:n
Sample (SA)

Legend

Mandatory table in hierarchy

Mandatory table not in hierarchy

Auxiliary table not in hierarchy
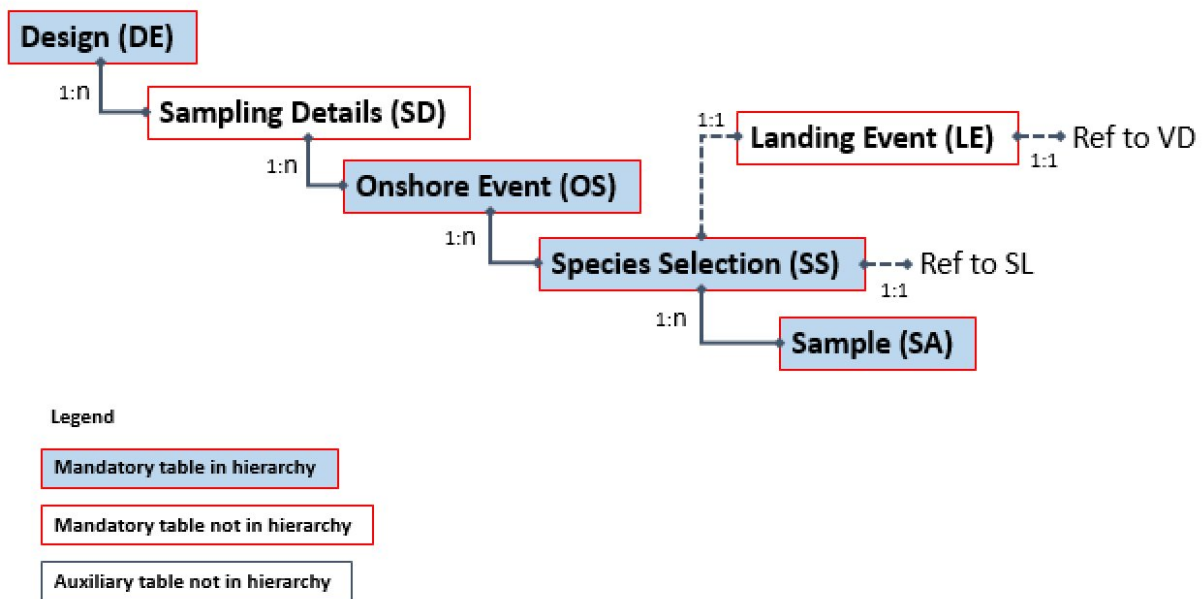
Example(s) described:

In the sampling design of Institute X, the person responsible for the sampling choose markets to visit, and times to visit them, at their own discretion. On 11-01-2016 observers went to market A and only one species from the list was present: Plaice. They sampled length-frequency as well as biological variables from that species. Within the sample there are 4 different size categories. In the next market*date (market B, 14-01-2016) again only one species was present and a sample of Sole was taken. In this sample, only length-frequency measures were taken. Within the sample there are 3 different size categories.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 8 Sampling where time (e.g., days, weeks) is the primary sampling unit and vessel is the secondary sampling unit

Hierarchy 8 is most used for on-shore sampling, but can be used for at-sea sampling as well.

Different schemes fit the hierarchy 8:

- Onshore sampling of landings where a time period is first selected (e.g., a set of weeks) for sampling to take place. Then, in each time period, a set of vessels is selected for sampling. One (or more) landing events from those vessels are sampled and in each one (or more) species are sampled. Information from the fishing trips, including start and end-dates may be available but generally only a posteriori (e.g., from logbooks): Selection takes place on the landing event and not on the fishing trip (which may register other landing events elsewhere or at a later occasion).
- Variants:
  a. Sampling carried out in ports or auctions
  b. Sampling done by observers or subcontractors
  c. Selection of different time units (e.g., day, month)
  d. Sampling all species in the landing or only species from pre-defined list(s)
  e. Different sampling strategies for the landings (e.g., stratification by commercial name/group, stratification by commercial size, etc.)

Upper hierarchy tables:

- Design (DE)
- Temporal Event (TE)
- Vessel Selection (VS)
- Landing Event (LE)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Fishing trip (FT) (optional)
- Species List Details (SL) (mandatory)
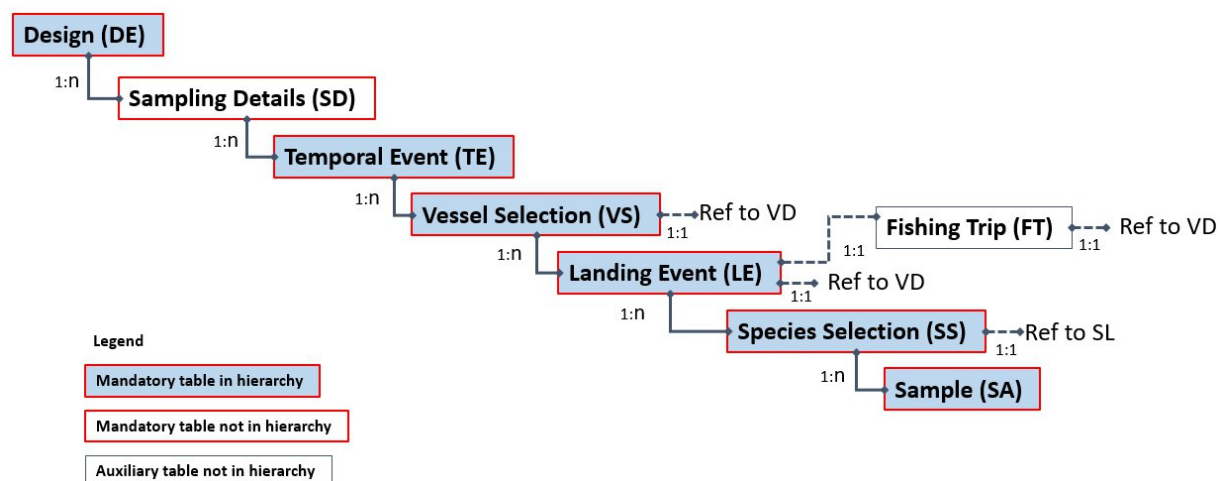
Example(s) described:

In the sampling design of a country (XYZ) an Institute (ANI) selected 8 weeks randomly each quarter for onshore sampling to take place. In each selected week, 4 vessels were selected for sampling. The list of vessels selected was sent to a subcontractor that was asked to sample one landing event from each of those vessels that landed that week. In each landing event only one species (cod) was sampled. Sampling was done by size category. 3 size categories were present: size 4, size 5 and size 7. From each 1 box was taken. All specimens in the sample were measured and aged.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Hierarchy 9 Sampling where location is the primary sampling unit, time is the secondary sampling unit and fish the tertiary sampling unit

Hierarchy 9 is most used for on-shore sampling.

Different schemes fit the hierarchy 9:

- On-shore sampling of landings, where a set of auctions (Location) are selected in the start of a period. Then, a set of dates are selected for sampling in each location (Temporal event). A set list of species are sampled within a location (Species Selection). Landing events are present but cannot be distinguished or are not of interest within the sampling. One or more boxes of fish are sampled from each species and size category (Sample). Each box of fish comes from one landing event or more landing events.
- Variants:
    a. Weight instead of boxes

    b. All locations are sampled within the period, so this works a bit like a reference fleet of locations
    c. One single location sampled within the period

Upper hierarchy tables:

- Design (DE)
- Location (LO)
- Temporal Event (TE)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Landing Event (LE) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)

Design (DE)

1:n

Sampling Details (SD)

1:n

Location (LO)

1:n

Temporal Event (TE)

1:n

Species Selection (SS) — 1:1 → Ref to SL

1:n

Sample (SA)

n:1

Landing Event (LE) — 1:1 → Ref to VD

Legend

Mandatory table in hierarchy

Mandatory table not in hierarchy

Auxiliary table not in hierarchy
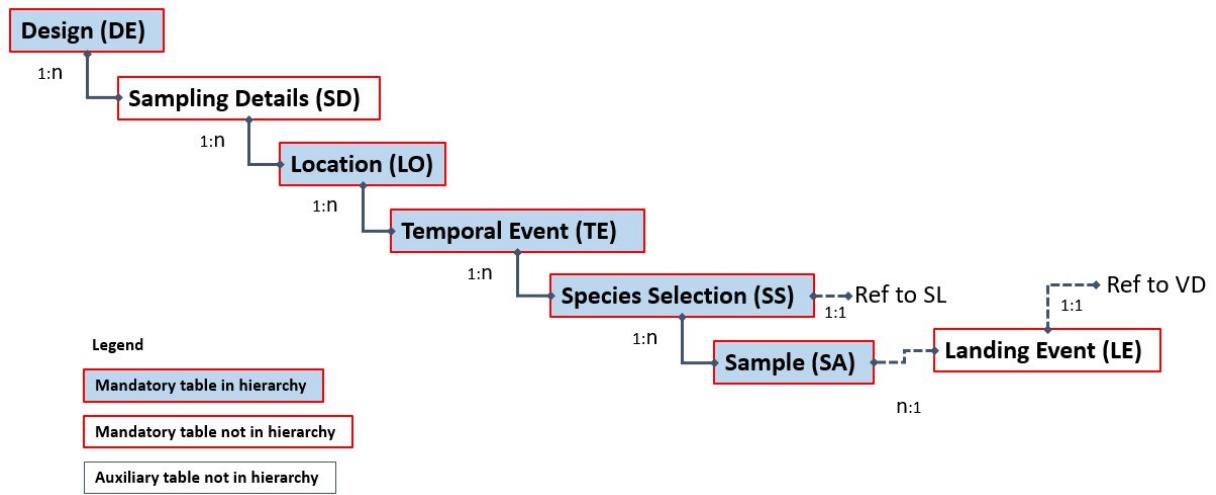
Hierarchy 10 Sampling where vessel is the primary sampling unit, time is the secondary sampling unit and trip is the tertiary sampling unit

Hierarchy 10 is most used for at-sea sampling, but can be used for on-shore sampling as well.

Different schemes fit the hierarchy 10:

- At-sea sampling, where a set of vessels (Vessel) are selected in the start of a period (e.g., a quarter). Then, one or more weeks (Temporal) are selected for sampling each vessel. In each week, contact is made and a trip is booked (Trip). In the trip, one or more hauls are sampled (Fishing Operation). A set of species is sampled in each haul (Species Selection) from the different catch categories. One or more boxes of fish are sampled from each selected species*size category.
- Variants:
    a. Vessels selected with UPWOR at quarterly level
    b. One or more weeks sampled per vessel
    c. Weeks only used to contact vessel with next trip selected
    d. Weight instead of boxes
- This hierarchy is particularly useful if a list of vessels are sampled without replacement at quarterly level

Upper hierarchy tables:

- Design (DE)
- Vessel (VS)
- Temporal Event (TE)
- Fishing Trip (FT)
- Fishing Operation (FO)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)
-

```
Design (DE)
   │
   │ 1:n
   └─── Sampling Details (SD)
            │
            │ 1:n
            └─── Vessel Selection(VS) ┄┄▶ Ref to VD
                     │                1:1
                     │ 1:n
                     └─── Temporal Event (TE)
                              │
                              │ 1:n
                              └─── Fishing Trip (FT) ┄┄▶ Ref to VD
                                       │             1:1
                                       │ 1:n
                                       └─── Fishing Operation (FO)
                                                │
                                                │ 1:n
                                                └─── Species Selection (SS) ┄┄▶ Ref to SL
                                                         │                  1:1
                                                         │ 1:n
                                                         └─── Sample (SA)
```

Legend

| Mandatory table in hierarchy |
| Mandatory table not in hierarchy |
| Auxiliary table not in hierarchy |

Hierarchy 11 is most used for on-shore sampling, but can be used for at-sea sampling as well.

Different schemes fit the hierarchy 11:

- On-shore sampling at landing ports or vendors where locations are selected first (e.g. in the beginning of a year). Each location is then sampled at random times (e.g. days). Within the time selected, some fishing trips are then sampled.
- Variants:

    a. Sampling at markets, buyers or vendors
- Note:
    a. This hierarchy is similar to hierarchy 6, the difference being that different types of probability sampling can be used for location and time and both can be specified. Use hierarchy 6 if you are using the same selection methods across space and time.

Upper hierarchy tables:

- Design (DE)
- Location (LO)
- Temporal Event (TE)
- Fishing Trip (FT)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Landing Event (LE) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)

Hierarchy 12 Sampling where location is the primary sampling unit, time is the secondary sampling unit and landing event the tertiary sampling unit

Hierarchy 12 is most used for on-shore sampling, but can be used for at-sea sampling as well.
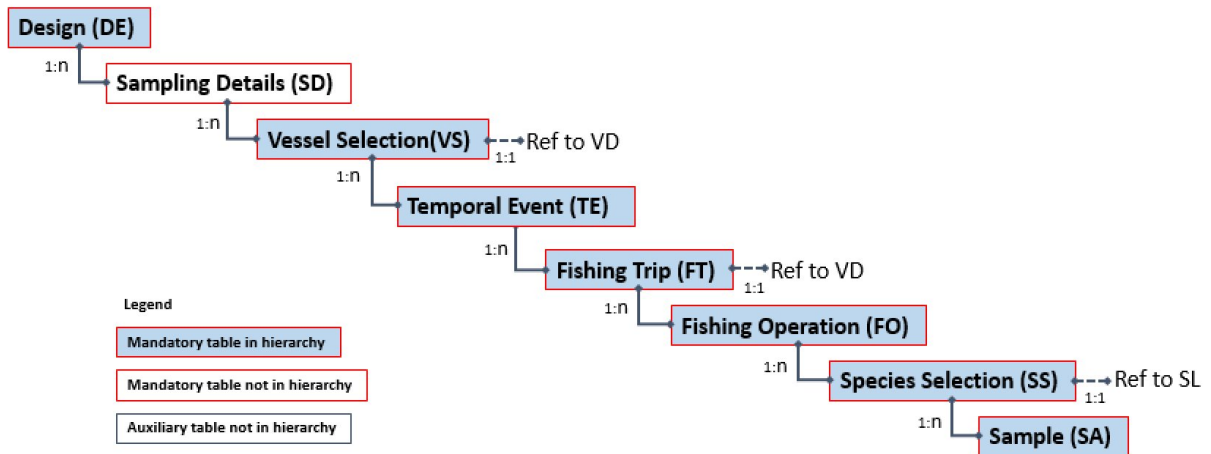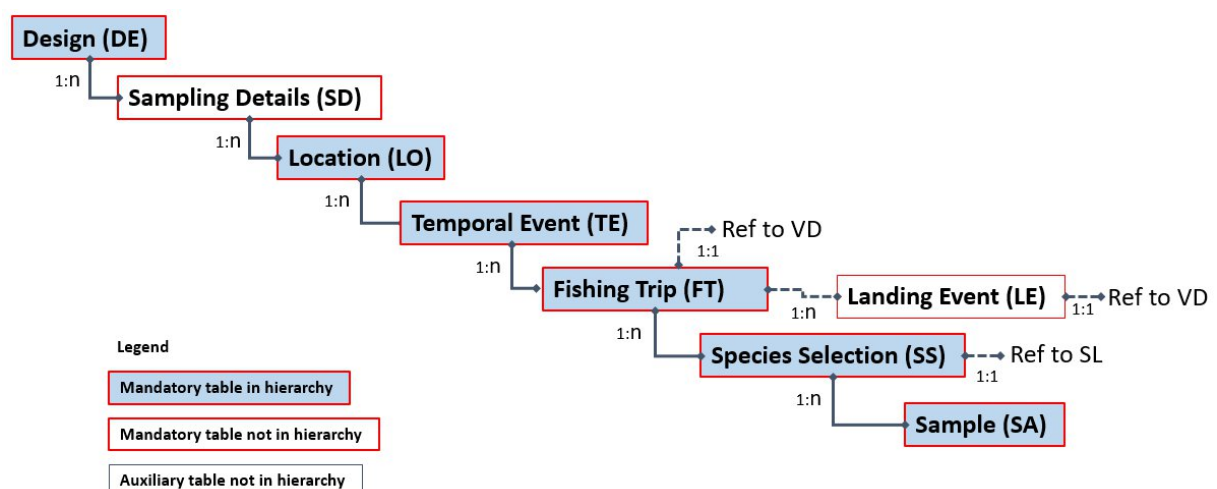
Different schemes fit the hierarchy 12:

- On-shore sampling landing ports where locations are selected first (e.g. in the beginning of a year). Each location is then sampled at random times (e.g. days). Within the time selected, some fishing landing events are then sampled. Data on the associated fishing trips is generally known, but not part of the selection procedure.

- This hierarchy is very similar to hierarchy 5, the difference being that instead of selecting an on-shore event (location*time), then this is split into first selecting a location and then selecting time (e.g. day). A bit like having a reference fleet of locations.
- Variants:
    a. Fishing trip is not known
    b. Sampling at markets, buyers or vendors
- Note:
    a. If you use equal-probability selection methods (SRSWR or SRSWOR) on both time and space components consider using hierarchy 5 instead.
    b. You can use this design for situations when you select first a set of times (e.g., days) and then sample one or more locations at each selected time

Upper hierarchy tables:

- Design (DE)
- Location (LO)
- Temporal Event (TE)
- Landing Event (LE)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Landing Event (LE) (mandatory)
- Vessel Details (VD) (mandatory)
- Species List Details (SL) (mandatory)

Design (DE)
1:n
Sampling Details (SD)
1:n
Location (LO)
1:n
Temporal Event (TE)
1:n
Landing Event (LE) ---> Ref to VD
1:1
Fishing Trip (FT)
1:1
1:n
Species Selection (SS) ---> Ref to SL
1:1
1:n
Sample (SA)

Legend
Mandatory table in hierarchy
Mandatory table not in hierarchy
Auxiliary table not in hierarchy

Hierarchy 13 Sampling where fishing operation is the primary sampling unit

Hierarchy 13 is most used for at-sea sampling, but can be used for on-shore sampling as well.
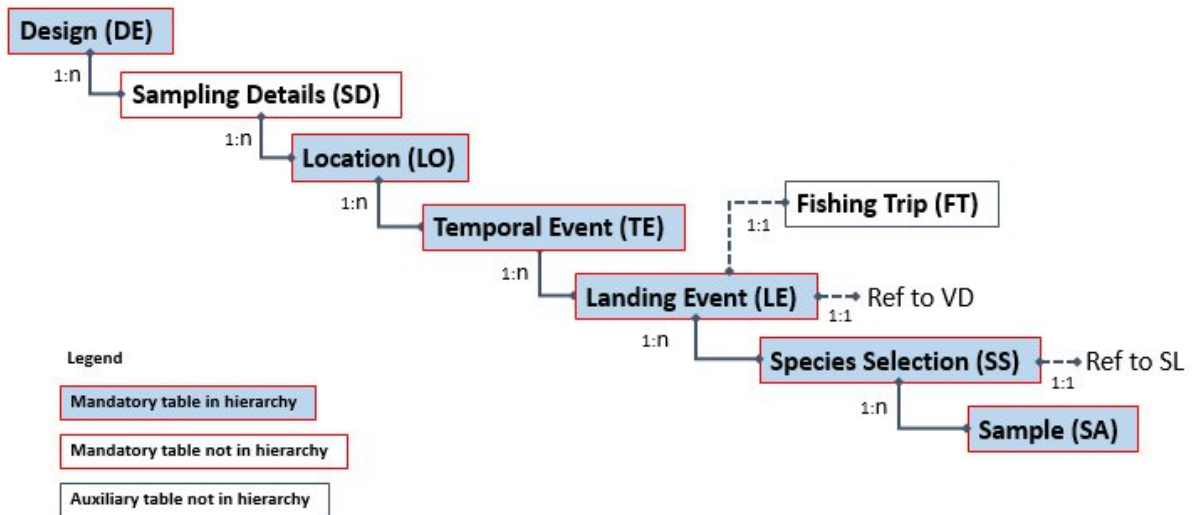
Different schemes fit the hierarchy 13:

- Sampling where individual fishing operations can be sampled directly, without any prior vessel or trip selection, for instance via real time reporting of fishing activity.

Upper hierarchy tables:

- Design (DE)
- Fishing Operation (FO)
- Species Selection (SS)
- Sample (SA)

Auxiliary tables:

- Sampling Details (SD) (mandatory)
- Fishing Trip (FT) (optional)
- Species List Details (SL) (mandatory)
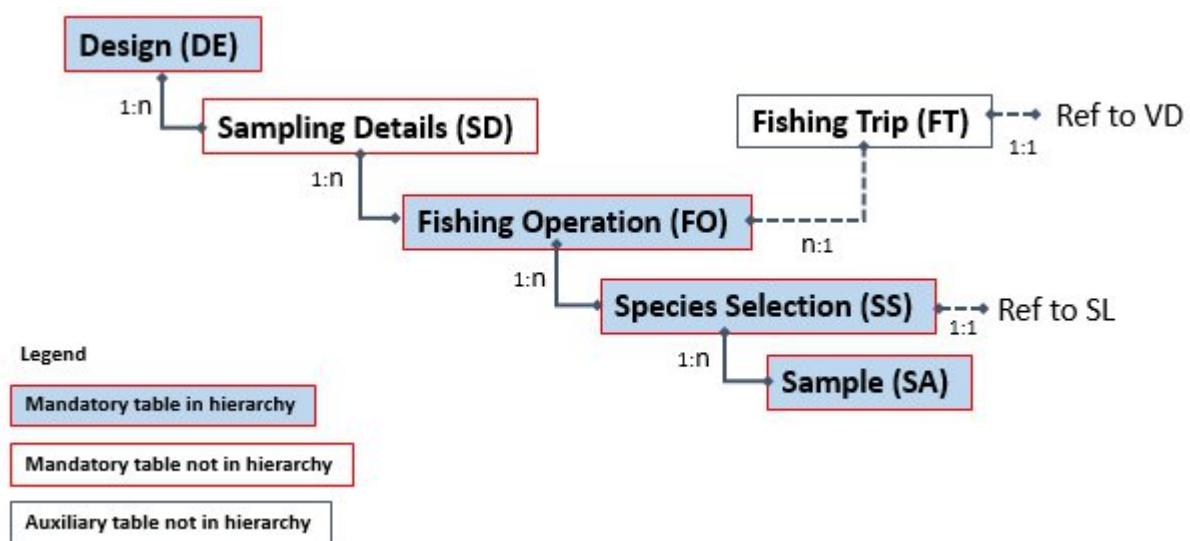
Lower Hierarchy A: Length stratified biological samples

Different schemes fit the lower hierarchy A:

- The sample of a specific species is length measured and stratified by length into 1 cm classes. A fixed number of specimens is biologically sampled from each class (e.g., 2 individuals from each length class). Length, weight, sex and age are determined for each of them.
- Variants:
    a. Different length classes (e.g., 1mm, 0.5cm, 2.5cm, etc.)
    b. Different sample sizes from each length class (e.g., proportional to the number of fish in the length class).

  c. Additional biological variables on each fish sampled (e.g., maturity, stock, age quality, etc.)

Upper hierarchy link:

- Sample (SA)

Lower hierarchy tables:

- Frequency Measure (FM)
- Biological Variable (BV)
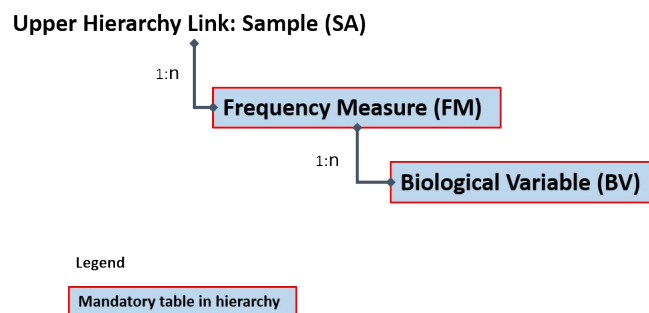
Example(s) described:

In a sample of species X, there are 18 fish. All fish are measured to the lowest cm and the sample is length stratified into 1cm length classes. The number of individuals in each length class is noted. Then, biological data is collected from up to 2 fish randomly selected from each length class. Data on length (in mm), weight (in g), age (in years), age quality and stock is available on each individual biologically sampled.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Lower Hierarchy B: Only length frequency data is taken from sample(s)/subsample(s)

Different schemes fit the lower hierarchy B:

- The sample of a specific species is length measured to the lowest cm. No further biological analyses are carried out on the specimens.
- Variants:
  a. Different length classes (e.g., 1mm, 0.5cm, 2.5cm, etc.)

Upper hierarchy link:

- Sample (SA)

Lower hierarchy tables:

- Frequency Measure (FM)

Upper Hierarchy Link: Sample (SA)

1:n

Frequency Measure (FM)

Legend

Mandatory table in hierarchy

Example:

In a sample of species X, there are 41 fish. All fish are measured to the lowest cm and data is organized into a 1-cm length frequency. No further biological data is collected from specimens.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Lower Hierarchy C: All individuals in the sample/subsample are biologically analyzed

Different schemes fit the lower hierarchy C:

- The sample/subsample of a specific species is entirely biologically analyzed.
- Variants:
    - a. Various biological variables collected on each fish sampled (e.g., maturity, stock, age quality, etc.)
    - b. Length is not among the variables sampled (e.g., sampling directly for ages)

Upper hierarchy link:

- Sample (SA)

Lower hierarchy tables:

- Biological Variable (BV)

Upper Hierarchy Link: Sample (SA)

1:n

Biological Variable (BV)
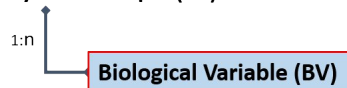
Legend

Mandatory table in hierarchy

Example:

In a sample of species X, there are 4 fish. All fish are biologically analysed with length, weight, age and sex determined.

For a filled-in example see spreadsheet "Examples of Hierarchies 1-8 A-D.xlsx".

Lower Hierarchy C: All individuals in the sample/subsample are biologically analyzed

Lower Hierarchy D: No length measurements or biological analyses

Different schemes fit the lower hierarchy D:

- Sample/subsample(s) of species that are weighed in bulk but not measured or biologically analysed (e.g., discards of some invertebrates; unsampled species from concurrent sampling of landing events).

Upper hierarchy link:

- Sample (SA)

Lower hierarchy tables:

- None

Example:

In a sample of species X no fish are measured or biologically analysed.

Annex: II RDBES

It has been recognised for many years that there is a need to have a new version of the RDB, which is the Regional DataBase and Estimation System development, RDBES, which also accommodate upload of statistical sampling information and statistical estimations.

It is important that the RDBES have only approved estimation methods and it is transparent regarding the processing and estimation of data. This is in the interest of the countries, ICES and the European Commission. The countries benefit from having one system where they can do the estimations. It is known that it is time consuming to do the complex estimations and there is risk of introducing errors, when estimating data for the ICES data call for the ICES stock assessment and advice. By using the RDBES, the RDBES will ensure high data quality and the user make sure the data is ready for estimation and select the right statistical estimation method. This will reduces the work load of the countries. The countries will also benefit from the common repository of all the countries' developed estimation methods, this should also reduce the work load of reinventing estimation methods in the countries in parallel. The Transparent Assessment Framework, TAF, will be used for developing estimation methods and executing the estimations, and the RDBES will also be storing the final estimation methods. When the TAF has been used for estimating data, ICES does not need to issue stock assessment data calls, because the ICES Expert Groups can get the estimated data from the TAF, which is needed for the stock assessment. This will reduce the numbers of data calls, which will benefit the countries. The RDBES will replace the RDB and InterCatch, and there have been a shift regarding combining the RDBES and the Transparent Assessment Framework, TAF, regarding the estimations – the broad work flow will be as follows.

- Before a stock will be included in the RDBES a data workshop, which covers several species, must be held. This will require the countries that are relevant for that stock to upload data in the new RDBES format. The experts at the workshop will then look at the data and the existing national estimation techniques to agree on the stock estimation techniques that should be used.
- Once the data workshop has been concluded then future data calls for that stock will require countries to have uploaded data to the RDBES by a deadline, rather than submitting raised estimates. The estimates will then be produced by running R scripts on the RDBES data.
- At the data workshop it should be agreed who will create the stock estimates. E.g. will each country create national estimates, which are then combined into stock assessments, or will a single person, in dialog with the national expert create the stock estimates.
- The techniques agreed in the data workshop will then be applied within the Transparent Assessment Framework (TAF) and the RDBES. If there are reasons why the agreed techniques should be modified in a particular year, then it is a national responsibility to inform the estimator(s).
- The stock estimates are then used as an input to the TAF Stock Assessment, this will be the second loop in TAF.

The two main end users are the RCGs and ICES, but data could also be downloaded/exported for data calls for other relevant end users e.g. STECF or ICCAT.

The figure below gives an overview of the flow of data in the RDBES and TAF.
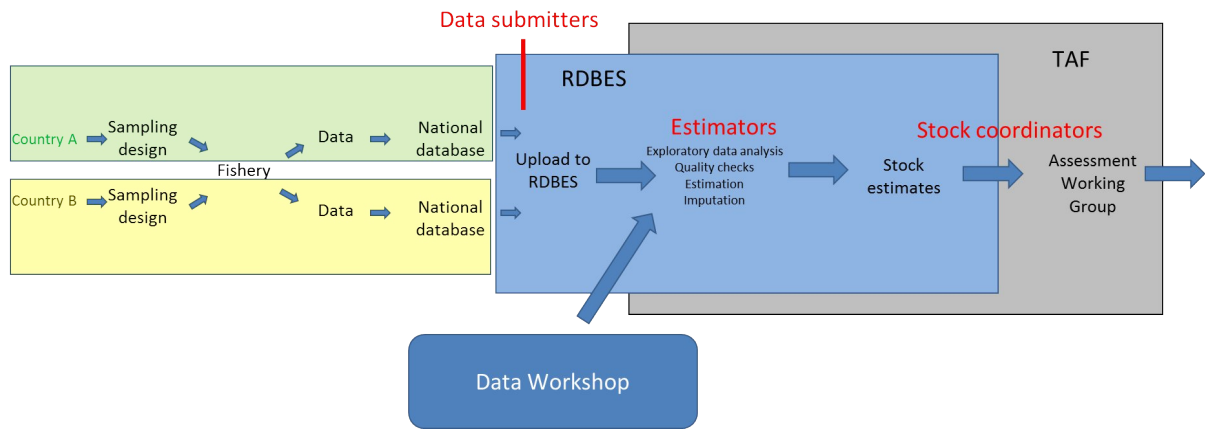
Annex III: Frequently asked questions

Question: What are sequence numbers, why are they in all selection tables and what is the difference between them and the unit names?

Identifying the number of unique units present in a given sample is needed for some estimation methods (e.g., obtaining Horvitz-Thompson estimates from data collected using with-replacement methods). Furthermore, knowing the order of unit selection in sampling events is required for some bootstrap variance-estimation methods. In the RDBES sampled units are identified by "unitNames" and the order of selection by "sequenceNumbers".

When sampling method is without replacement (e.g., SRSWOR), a unit can only be selected for sampling once. This results in sequence numbers and unit names being unique within a sample. However, when sampling is done with replacement (e.g., SRSWR), some units can be selected two (or more) times generating duplications in unit names. Still, those units were selected in a certain order (or sequence) that clearly distinguishes them. The importance of the order of unit selection is not readily apparent in some estimates frequently calculated (e.g., means and variances of SRSWR); but becomes fundamental when one wants to consider bootstrapping of more complex designs involving, e.g., unequal probability sampling.

The following table shows conceptually the functioning of variables "sequenceNumber" and "unitName" in an RBDES table where data was collected by Simple Random Sampling With Replacement (SRSWR). In such situation, because the sampling was done With Replacement, duplicates are acceptable in variable unitName. In the present case, unit "A" was sampled twice (present twice in unitName); and sequenceNumber orders the two events (e.g., in time) allowing their distinction.

sequenceNumber	Method	unitName
1	SRSWR	A
2	SRSWR	B
3	SRSWR	C
4	SRSWR	A

The following table shows conceptually the functioning of variables "sequenceNumber" and "unitName" in an RDBES table where data was collected by Unequal Probability Sampling Without Replacement (UPSWOR). In such situation, because the sampling was done Without Replacement, no unit can be sampled more than once, so both unitNames and sequenceNumbers are unique.

sequenceNumber	Method	unitName
1	UPSWOR	A
2	UPSWOR	B
3	UPSWOR	C
4	UPSWOR	D

Question: Many vessels can be unequivocally identified by a combination of flag country and length and/or power and/or tonnage. Is it possible to maintain the anonymity of vessels in the RDBES data model?

Answer. Yes. Those characteristics are present in the Vessel Details table and are optional. In the rarer cases where length class unequivocally identifies the vessel, that variable can be set to "unknown".

Question: Can the RDBES store data from sampling programmes, where the sampling units are, e.g., boxes of commercial designation that may hold multiple species (e.g., anglerfishes; rays)?

Answer. Yes. If sampling is done on a, e.g., subset commercial designations, the target commercial designations can be stated in a species list (Variable 6: SSspeciesListID in the Species Selection Table). Then, the boxes of the commercial designation can be entered as rows in the Sample table (Variable 8: SAcommercialSpecies) and subsamples of each box entered for each individual species present.

Question: Can the RDBES account for the possibility of fishing area being a strata or a domain of a fishing trip observed at sea?

Answer: If the sampling programme involves stratification of the fishing operations of each fishing trip by area then 'FOnumberTotal' and 'FOnumberSampled' must reflect the total number of fishing operations made and sampled from each area visited by the fishing trip. If the sampling programme does not involve stratification by area, e.g., if a systematic sample is taken from all fishing operations in the fishing trip and the area is only recorded, the 'FOnumberTotal' and 'FOnumberSampled' refers to the fishing trip. If e.g. total number of fishing operations are needed per area for e.g. ratio estimation, then the Fishing Operation table needs to be populated with all fishing operations (sampled and not-sampled).

Question: Can the RDBES store data from sampling programs at-sea where e.g. length measurements (Lower Hierarchy B) are collected per fishing operation, but fish for individual measurements (Lower Hierarchy C) are collected throughout the fishing trip and are not linked to a specific fishing operation?

Answer: Yes. For the fish collected per fishing operation for length measurements: *Fishing Operation table: 'FOaggregationLevel='H'*. Species Selection table: Filled with appropriated information. Sample table: 'SAnumberTotal', 'SAnumberSampled', 'SAsampProb' and 'SAselectionMethod' are filled with the appropriated information and 'SAlowerHierarchy='B'. For fish collected for individual measurements throughout the fishing trip: *Fishing operation table: 'FOaggregationLeve'='T'*. Species Selection table: Filled with appropriated information. Sample table: 'SAnumberTotal', 'SAnumberSampled', 'SAsampProb' and 'SAselectionMethod' are left empty and 'SAlowerHierarchy'='C'.

Question: If in an onboard trip the weights of some species are registered at haul level with the weights of other species being registered at trip level, how can this information be entered in the RDBES?
Answer: In this case, the FO table should contain the rows of individual hauls (with FOaggregationLevel'='H') and one additional row corresponding to the whole trip (with FOaggregationLevel'='T'). Each haul can have entries in the Species Selection, SS, table with corresponding SS lists used at haul level, e.g., for landings and discards; and the general trip row will also have such lines with the SS table carrying the SS lists recorded at trip level.

Question: If in an onboard trip the lengths are collected at haul level and the ages sampled at trip level, how can this information be entered in the RDBES?

Answer: In this case, the FO table should contain the rows of individual hauls (with FOaggregationLevel'='H') and an extra row for the trip (FOaggregationLevel'='T'). The "H" rows will be connected to a SA row with lower hierarchy B (lengths). The "T" will be connected to a SA row with lower hierarchy C (ages). In the SA table the weights should only be present in the Haul samples.

Question: Can unequal probability (e.g., proportional to size) be specified in the RDBES?

Answer: Yes. You can do that by specifying "Unequal Probability Sampling With Replacement (UPSWR)" or "Unequal Probability Sampling Without Replacement (UPSWOR)" in selectionMethod and providing the inclusion probability of each element included in the sample in sampProb variable. See example below.

Example: Simplified example of a sample where elements were sampled with replacement and unequal probability proportional to size. A sample of vessels (n = 5)) was taken with replacement from a population of N=10 vessels. The probabilities of inclusion of the elements of the population were: 1001 = 1002 = 0.3; 1003 to 1010 = .05. Note the probabilities of inclusion must be provided in sampProb.

VDid	VSstratum	VSselectionMethod	VSnumberTotal	VSnumberSampled	VSsampProb
1	None	Unequal probability with replacement	10	5	0.3
2	None	Unequal probability with replacement	10	5	0.3
3	None	Unequal probability with replacement	10	5	0.3
4	None	Unequal probability with replacement	10	5	0.05
5	None	Unequal probability with replacement	10	5	0.3

Question: What is the difference between selecting a Fishing Trip (FT) and selecting a Landing Event (LE)?:

Answer: A Fishing Trip is defined as a vessel operating between departure and arrival date and times; A Landing event is defined as a vessel unloading/selling some catch (could be part of the catch; could be catch from several trips) in a specific port/market. The two concepts have distinct usages during onshore sampling. In most onshore sampling situations observers sample a vessel's landings but *do not know or are not sure* if what they are sampling represents the whole landings from a trip or if, e.g., the vessel still has some fish onboard to be landed elsewhere *or* has landed elsewhere before *or* has pooled landings from several previous fishing trips that it is selling all at once in that market-day. Additional cases exist when catch of a vessel is landed over several days in the same port but only one of those days do observers sample. In all such cases it is generally difficult to warrant that the landings being sampled are fully representative of a specific trip. Such difficulty is reflected in the RDBES by stating what is actually being sampled are the landings from one (or more) *Landing Events* and not from one (or more) *Fishing Trips*. Note: For practical purposes a *Landing event* can be considered as the landings made by a vessel in a specific place and time while the *landings of a Fishing Trip* are the landings of a specific fishing trip in all places and times.

The decision as to whether you are sampling Landing Events or Fishing Trips will come down to your sampling scheme – is it specifically designed to always sample entire trips? If not, then you should probably use a sampling hierarchy that includes Landing Event.

Question: what is the difference between Hierarchy 6 and Hierarchy 7? Why does hierarchy 6 contain FO table if FO is not strictly part of the hierarchy?

Hierarchy 6 differs from Hierarchy 7 on the level of confidence existing on some characteristics of the samples and trip they came from. Hierarchy 6 should only be used in situations where trip information is well known and where samples collected can be unequivocally assigned to a certain area, gear and trip. This is only possible when trips involve a single gear and fishing area, and it is known that catch does not remain on-board vessels when they land (e.g., small-scale fisheries); Conversely, in Hierarchy 7 – a much more common alternative - a landing event may come from a trip that fished in more than one area or with more than one gear; and it is assumed that landing events samples may only partially represent the catch retained during a single trip (some catch may have been retained on board and not landed; transhipment may have occurred ahead of landing).

Hierarchy 6 contains FO table (and requires it set to "T" aggregation level) so that samples collected under this hierarchy have characteristics like area and gear unequivocally associated to them. Note that contrary to Hierarchy 7 area reported under hierarchy 6 is not dominant.

Question: why are area, gear, and some other auxiliary variables defined as "Dominant" in LandingEvent table?

It is not infrequent that fishing trips of medium and large-scale vessels span more than one area/rectangle or use more than one gear. When samples are collected at-sea after individual fishing operations, auxiliary information (e.g., area and gear characteristics) associated to individual samples is exactly known. This situation contrasts with onshore sampling of Landing Events that may contain specimens caught in multiple fishing operations. Such situation leads to difficulties in the determination of e.g. the area or gear that caught the individual fish, which limits some uses of the data [e.g., in stock discrimination or selectivity-related studies]. The RDBES accommodates and highlights these uncertainties by defining area, gear and other auxiliary variables present in the Landing Event (LE) table as "Dominant"; and area, gear and other auxiliary variables present in the Sample table (SA) as exact (or otherwise not filled in).

Example: Simplified example of area dominance in Landing Events and Samples. In LEid 1 an area is assigned to the Landing Event (LEid 1) but the exact area of the samples is not known, and therefore not filled in. In LEid 2 a dominant area is assigned to the Landing Event and it is known that the samples come from that area (e.g. VMS indicates only 27.3.d.28 was fished). LEid 3 represents a mixed situation – where the dominant area of the landing event is known but only partial information exists at sample level

View Landing Evernt (LE)

LEid	Area (M)
1	27.3.d.28
2	27.3.d.28

View Sample (SA)

LEid	SAid	Weight	Area (O)
1	1	100	
1	2	20	
2	3	150	27.3.d.28
2	4	50	27.3.d.28
3	5	100	27.3.d.28
3	6	50	27.3.d.27
3	7	20	

Example: Simplified example gear dominance in Landing Events and Samples. In LEid 1 a dominant gear is assigned to the landing (GNS) but the exact gear of individual samples is not known (not filled in). In LEid 2 a dominant gear is assigned to the landing (GNS) and the exact gear of individual trips is also known (only 1 GNS was used in the trip). In LEid 3 a dominant gear is assigned to the landing (GNS) and exact gear of individual samples is known (GNS, LLS, FPO) for all samples but one (SAid 8).

View Landing Event (LE)

LEid	Gear (M)
1	GNS
2	GNS
3	GNS

View Sample (SA)

LEid	Said	Weight	Gear (O)
1	1	100	
1	2	20	
2	3	150	GNS
2	4	50	GNS
3	5	100	GNS
3	6	50	LLS
3	7	10	FPO
3	8	10	

Question: In previous years my sampling protocol did not specify a species list (or it was too general, e.g., *commercial fish*). How should I specify this in the RDBES?:

These type of situations is quite common: e.g., in a specific programme observers were instructed to record all commercial species and they are confident that they always recorded the main ICES stocks but there is no consensus if they always screened the samples for some stocks which data only recently is being requested (e.g., a specific ray or cephalopod). How then to assign a species list to previous samples? The guideline is that only the species that one is confident were always on the protocol should be on the species list (i.e., only put on the list the species to which 0s can confidently assigned when they were absent). All other species, which one is not sure of, can and should still be uploaded to the RDBES but not be included in the list. By being outside the list they will be automatically flagged as uncertain.

Question: I have a species list but I found that in a group of trips my observers made a dubious interpretation (e.g., some groups were not consistently recorded)? Can I indicate this in the RDBES?

Yes. If for some reason in a specific set of trips the protocol was interpreted differently, you can declare for that subset of trips a special species list that includes only the subset of species consistently recorded in them. For the remainder, you can state the original *de facto* species list in your protocols.

Question: Is the RDBES prepared to receive data from marine litter?

Not yet. ICES has a EG dealing directly with marine litter (WGML) and litter data can in principle also be included in the RDBES in a way similar to observations from fish and by-catch. However this is beyond the current scope of RDBES. Albeit it is theoretically possible to obtain estimates of marine litter from commercial samples (e.g., from at-sea observations or analysis fish stomachs), few countries are known to collect these data and there is no routine request to ICES for estimates of marine litter from commercial fisheries.

Question: If some species are transhipped during an observed trip, how can I declare them in the RDBES?

With regards to transhipments a decision was made not to include them as catchCategory in the sample table (SA) as users can declare it as "Reason for Not Sampling" in other tables - e.g., if a specific part the catch was not sampled because of transshipment (e.g., a species) you can declare it in the Species Selection Table (e.g., under "SSreasonNotSampled"); If a non-sorted part of the catch was not sampled and you know its weight, you can declare it by stratifying the SA table with a non-sampled strata and then declare what happened under "SAreasonNotSampledFM".

Question: How are SSObservationActivityType "Sorting", "Pre-sorting", "Drop-outs" and "Slipping" defined in the different fisheries and vessel types?

WGBYC and WGCATCH are working on operational definitions for these terms in the different fisheries and types of vessels. Not all SSObservationActivityType will apply to all fisheries and vessels with some distinctions being necessary (e.g., for small-sized vessels). For purposes of the 2020 RDBES data call consider:

- Slipping – Intentional release of fish from the fishing gear before that gear is fully brought on board the fishing vessel (e.g., opening of a purse-seine in the water to release part of the catch)

- Drop-outs – Non-intentional release of catch from the fishing gear before that gear is fully brought on board the fishing vessel (e.g., specimens that loose themselves from hooks or meshes and fall back into the water as gear is being hauled onboard)

- Pre-sorting – catch left on deck and/or released to sea immediately after the gear is brought onboard (e.g., when cod-end is emptied in trawlers; as the hauling drum is operated in gillnetters and longliners)

- Sorting – sorting of the catch into landings and discards on deck on conveyor belt, sorting platform or on the vessel deck itself; storage of the catch unsorted (e.g., large pelagic trawlers and purse-seiners)

Question: We often sample a cluster of fish e.g. a box or a basket, but in some cases, we actually select a subsample of fish in an unclustered fashion e.g. selecting single fish in a randomly way throughout a haul. How can this be declared in the sample table (SA)?

Answer: Since each fish is considered a unit sampled, then each fish could get a row in the sample table. This would be a very tiresome way of recording this, so it is possible to group the fish into one row and the using SAunitType = 'individual units' to declare that these units are unclustered.

Question: A common approach to sampling fish from e.g. a haul at-sea would be to take an amount of fish (kg) out of the total amount of fish (kg). Such sampling protocols does not identify how the haul is divided into sampling units, unlike sampling via e.g. baskets or time, so how should this be handled in the sample table (SA) in respect to selection method, selection probability (p), inclusion probability (π) and Estimation of CV's? A variation of this would be that the sampling protocol states that only a single unit (e.g. basket of fish) should be subsampled, or it actually states that several units should be subsampled, but only the amount of fish (kg) in the subsample and the total amount of fish (kg) is recorded, no information about basket sampled and basket total are recorded.

Answer: Kg subsampled/kg total is not an inclusion probability but just a proportion. The inclusion probability is the probability that a sampling unit (here the kg subsample) is included in the final sample of n sampling units.

Say, the haul consists of a total weight TW and sample weight SW, where SW ≤ TW.

There are two possible ways to handle the issue:

1. The haul consists of approximately N=$\lfloor$TW/SW$\rfloor$ (where $\lfloor a \rfloor$ is the floor function, the largest integer smaller than a) possible sampling units. One unit, i.e. n=1, was selected at random and without replacement (SRSWOR). Hence, the selection and inclusion probabilities are the same and equal to 1/N. There is no such thing as a joint inclusion probability since there is only one unit was sampled and the variance is 0. This is probably the more accurate approach in the sense that it reflects the actual sampling design, but there will not be any way of measuring sampling variability. So for the example if 100 kg was sampled out of 547 kg, then N = floor(547/100) = 5 and π_i = 1/5 = 0.2 and p_i = 1/5 = 0.2. In this case the fish are considered a cluster.

2. The haul, TW, consists of N fish, where N is unknown. N would be estimated using either (a) length-weight relationships and proportions at length from the sample, (b) by dividing the total weight TW by the mean weight of a fish or (c) N=$\lfloor$TW/SW$\rfloor$*n. Now, SW consists of n sampled fish where n is known but a random number. What is often done when one wants to determine both selection and inclusion probabilities of units sampled is to assume (a) the number of fish to be sampled, n, is a fixed number (i.e. you planned it) and (b) the fish were sampled at random and with replacement (SRSWR). If these 2 assumptions are reasonable, then the selection probability is p_i = 1/N, the inclusion probability is π_i = 1-(1-1/N)^n, and the pairwise joint inclusion probability is given by π_{ij} = 1-2(1-1/N)^n+(1-2/N)^n. In this approach, all of the numbers used in the calculations of the probabilities are approximate but you can also estimate a sampling variance for the estimated quantities of interest. In this case the fish are considered to be unclustered.

Answer adopted from ICES, 2020[4]

Question: I have selected a fixed set of units that I routinely sample (e.g., always the same ports or weeks of the year). How should these be entered in the RDBES?

This topic is scheduled to be discussed by a WGCATCH intersessional subgroup during 2020. For immediate purposes of the 2020 RDBES data call the core group provisional advice on this issue is that total number of units and total number of sampled units are set equal, method set to "NPEJ" and probability of inclusion set to 1. See the following examples:

- Country A routinely samples 7 ports out of 20 available. In the location table (LO) it should report 7 rows where totalN=7 ports, n=7 ports, method = NPEJ, and probInclusion=1
- Country B routinely samples 40 weeks out of 52 available in the year. In the temporal table (TE) it should report 40 rows where totalN=40 weeks, n=40 weeks, method = NPEJ, and probInclusion=1

Question: When in an onboard trip the length frequencies may be collected at haul level but the sampling goals in terms of number of ages that observers aim to collect per length-class is set at trip level. How should this information be entered in the RDBES?

The situation described is one of quota-sampling for ages at trip-level and will requires careful documentation so it can be correctly signaled to those carrying out estimation. Two variants are known to occur:

[4] ICES. 2020. Workshop on Estimation with the RDBES data model (WKRDB-EST; outputs from 2019 meeting). ICES Scientific Reports. 2:5. 106 pp. http://doi.org/10.17895/ices.pub.5954

1. The information about which hauls the individual fish are coming from is not recorded and stored, so it is not possible to link the individual fish to a specific haul
2. The information about which hauls the individual fish are coming from is recorded and stored, so it is possible to link the individual fish to a specific haul

In the case of situation 1, the FO table should contain the rows of individual hauls (with FOaggregationLevel'='H') and an extra row for the trip (FOaggregationLevel'='T'). The "H" rows will be connected to a SA row with lower hierarchy B (= only length frequency) where the length-frequencies collected in each haul can be recorded. The "T" row will be connected to a SA row with lower hierarchy C (= only bio sampling). In the SA table there will be weights in all samples that respect to the individuals sampled per haul for length frequency (hierarchy B) and in the trip for biological sampling (hierarchy C). It is important that in the SA table the selection method used for the biological samples (hierarchy C) is "quota sampling" so estimators know that quota sampling took place at trip level.

In the case of situation 2, the FO table should only contain the rows of individual hauls (with FOaggregationLevel'='H'). Those "H" rows will be connected to a SA row with lower hierarchy B (= only length frequency) where the length-frequencies collected in each haul are to be recorded. With regards to ages, two things can happen. In both cases Hierarchy C is used, sampling method is set to "quota sampling", and the weights correspond to the individuals sampled:

2a. The specimens sampled were taken from the same sample as the length frequency: one subsample is placed under the length-frequency sample.

2b. The specimens sampled were taken from an additional sample other than the sample used for length frequency: one additional sample is open.

Annex IV: Incidental bycatch and slipping

Situation 1: In a trawler, observers take their volume samples from the conveyor belt (separate samples on discards and landings) but do not observe other areas of the boat, e.g., they are not present at cod-end opening. All species in the samples are identified and weighed (species list B = "all species in sample")
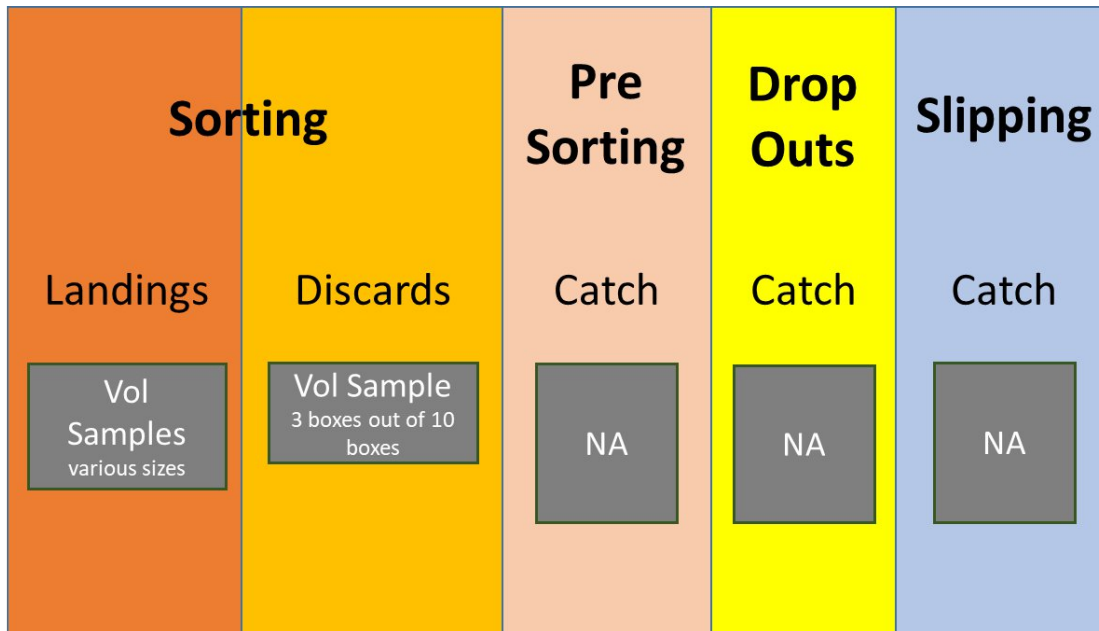
FO Table

FOid	...	FOobservationCode	Observation code explained
1	...	So	only Sorting observed
...	...	...	...

SS Table

SSid	Foid	...	SSobservationActivity	SScatchFraction	SSobservationType	SSsppListName
1	1	...	SORT	LAND	1	B
2	1	...	SORT	DIS	1	B
...						

Situation 2: In a gill-netter, observers screen visually the sorting of the catch for approximately 260 min out of the 360 minutes the operation takes. They also spend some time (60 min) on the edge of the boat observing the net being hauled and counting specimens that fall back into the water or in the deck close to the drum as the net is being hauled. They make their observations with regards to the species in a list A.
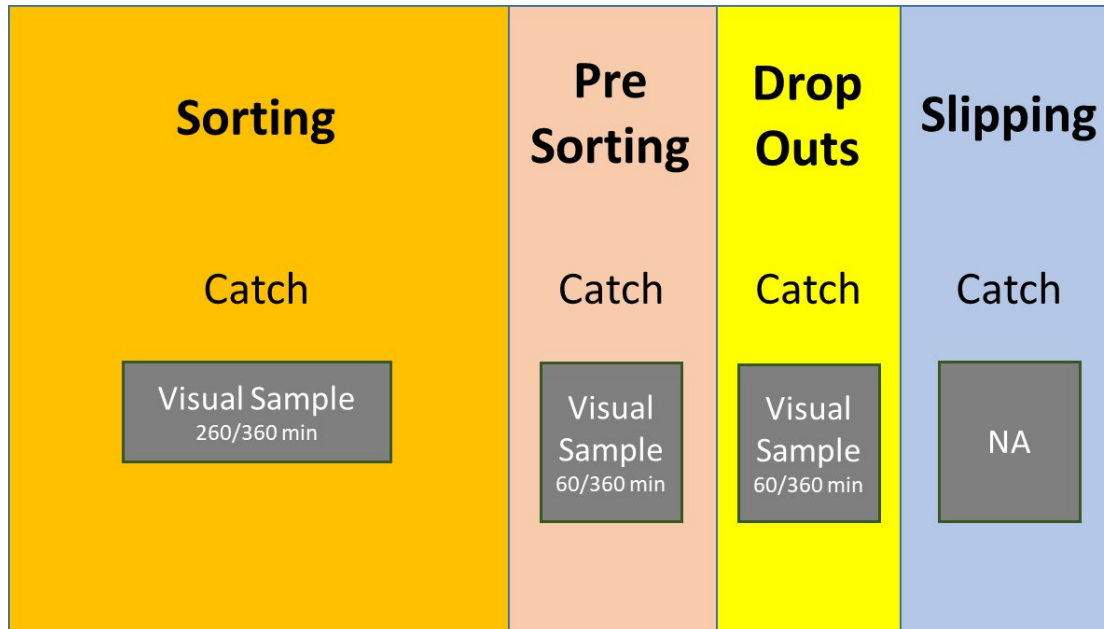
FO Table

FOid	...	FOobservationCode	Observation code explained
1	...	DrPrSo	Drop-outs, Presorting and Sorting observed
2	...	DrPrSo	Drop-outs, Presorting and Sorting observed
...	...	...	...

SS Table

SSid	Foid	...	SSobservationActivity	SScatchFraction	SSobservationType	SSsppListName
1	1	...	DROP	CATCH	2	A
2	1	...	PRESORT	CATCH	2	A
3	1	...	SORT	CATCH	2	A
4	2	...	DROP	CATCH	2	A
5	2	...	PRESORT	CATCH	2	A
6	2	...	SORT	CATCH	2	A
...	...	...	...	...	...	...

Situation 3: In a purse-seiner, observers visually screen the entire closure of the net and the hauling of the catch onboard with regards to the occurrence (or not) of individuals from a short-list of incidental by-catch species (List A). Throughout they pay attention and record the occurrence of slipping and drop-outs. Then they take volume samples from both landings and discards where they weight all species found (list B).

FO Table

FOid	...	FOobservationCode	Observation code explained
1	...	SlDrPrSo	Slipping, Drop-outs, Presorting and Sorting observed
...	...	...	...

SS Table

SSid	Foid	...	SSobservationActivity	SScatchFraction	SSobservationType	SSsppListName
1	1	...	SLIP	CATCH	2	A
2	1	...	DROP	CATCH	2	A
3	1	...	PRESORT	CATCH	2	A
2	1	...	SORT	LAND	1	B
3	1	...	SORT	DIS	1	B
...	...	...	...	...	...	...

Situation 4: Similar to situation 3 but observers also visually screen the occurrence (or not) of individuals from a short-list of incidental by-catch species (List A) during the sorting of the catch.

FO Table

FOid	...	FOobservationCode	Observation code explained
1	...	SlDrPrSo	Slipping, Drop-outs, Presorting and Sorting observed
...	...	...	...

SS Table

SSid	Foid	...	SSobservationActivity	SScatchFraction	SSobservationType	SSsppListName
1	1	...	SLIP	CATCH	2	A
2	1	...	DROP	CATCH	2	A
3	1	...	PRESORT	CATCH	2	A
4	1	...	SORT	CATCH	2	A
5	1	...	SORT	LAND	1	B
6	1	...	SORT	DIS	1	B
...	...	...	...	...	...	...

Annex V: Algorithms for the calculation of both inclusion and selection probabilities

By Mary Christman (ICES, 2020[5])

Probabilities at a given level in a sampling scheme

Type of Sampling	Selection Probability	Inclusion Probability (π_i)
SRSWR	$\dfrac{1}{N}$	$1 - \left(1 - \dfrac{1}{N}\right)^n$
SRSWOR	See below	$\dfrac{n}{N}$
UPSWR	p_i	$1 - \left(1 - p_i\right)^n$
UPSWOR	See below	$\displaystyle\sum_{k=1}^{n} p_i(k)$

Notation: let

N = number of elements in the population, i.e. the population "size",

n = the number of elements to be sampled, i.e. the sample size,

WITH REPLACEMENT

p_i = the probability of selection with replacement of the ith element in the population, $i = 1,...,N$, where $\sum_i p_i = 1$

WITHOUT REPLACEMENT

k = the order in which the ith element is sampled from the population when sampling is without replacement

$p_{i_j}(k)$ = probability that the ith unit is selected on the jth draw when k draws have occurred, $j \leq k$. For example, $p_{i_1}(2)$ is the probability that the ith unit was selected on the 1st draw (hence the i_1) when $k = 2$ draws have occurred so far

$$\frac{p_{i_k}(k)}{1 - p_{i_1}(k) - p_{i_2}(k) - ... - p_{i_{(k-1)}}(k)} = \frac{p_{i_j}(k)}{1 - \sum_{m=1}^{(k-1)} p_{i_m}(k)} = \text{conditional probability that the } i_j^{th} \text{ element is selected on}$$

the kth draw given that the $i_m^{th}, m = 1,2,...,(j-1)$ elements where selected prior to the kth draw, e.g. $\dfrac{p_{i_2}(2)}{1 - p_{i_1}(2)}$ is the probability that the i_2^{th} element is selected on the 2nd draw given that the i_1^{th} element was selected on the first draw.

---

[5] ICES. 2020. Workshop on Estimation with the RDBES data model (WKRDB-EST; outputs from 2019 meeting). ICES Scientific Reports. 2:5. 106 pp. http://doi.org/10.17895/ices.pub.5954

$$p_i(k) = \sum_{(k-1),i} p_{i_1}(1) \times \frac{p_{i_2}(2)}{1-p_{i_1}(2)} \times \frac{p_{i_3}(3)}{1-p_{i_1}(3)-p_{i_2}(3)} \times \cdots \times \frac{p_{i_k}(k)}{1-p_{i_1}(k)-p_{i_2}(k)-\ldots-p_{i_{(k-1)}}(k)} = \text{the unconditional}$$

probability that the ith element is selected at the kth draw where the summation notation indicates that the sum is over all possible sets of $(i_1,i_2,\ldots,i_{(k-1)})$ where the i_m are different integers between 1 and N and none equal i. For each k, $\sum_i p_i(k) = 1$

Example calculations using SRSWOR

1) The probability of the ith element being selected on the first draw is

$$p_i(1) = \frac{1}{N}$$

and the probability of not being selected in the first draw is

$$p_i^C(1) = 1 - \frac{1}{N}.$$

2) The probability that the ith element is not selected on the first draw but is selected on the second draw is

$$p_i(2) = \Pr\left(i^{th} \notin 1^{st} \ draw\right) \times \Pr\left(i^{th} \ in \ 2^{nd} \ draw \middle| \notin 1^{st} draw\right) = \left(1 - \frac{1}{N}\right)\frac{1}{N-1} = \frac{1}{N}$$

3) The probability that the ith element is selected on the kth draw $(2 < k \leq n < N)$ is

$$p_i(k) = \Pr\left(i^{th} \ not \ in \ 1^{st} \ k-1 \ draws\right) \times \Pr\left(i^{th} \ in \ k^{th} \ draw \middle| not \ in \ (k-1) \ draws\right)$$

$$= \left(1 - \frac{1}{N}\right)\left(1 - \frac{1}{N-1}\right)\left(1 - \frac{1}{N-2}\right)\cdots\left(1 - \frac{1}{N-k-1}\right)\left(\frac{1}{N-k+1}\right) = \frac{1}{(N-k+1)}\prod_{j=0}^{k-2}\left(1 - \frac{1}{N-j}\right)$$

$$= \frac{1}{N}$$

4) The inclusion probability for the ith element is

$$\sum_{k=1}^{n} p_i(k) = \sum_{k=1}^{n} \frac{1}{N} = \frac{n}{N}$$

Note that the probabilities are identical only because every element has the same chance of being selected at each draw, so the equations for the probabilities at each draw are simplified. This is not the case for unequal probabilities.

Annex VI: Examples of the use of Stratification and Clustering

The RDBES provides the possibility of declaring stratification and/or clustering at each level of the sampling hierarchy.

Stratified sampling considers the sampling units of a sampling level as mutually exclusive sets that are all sampled independently with a sometimes stratum-specific sampling effort. When stratified sampling is used it is fundamental for design-based estimation that every stratum has been sampled because separate estimates are to be calculated for each stratum that will later be added together to obtain the total or mean for that sampling level. A simple example of stratification is the spatial stratification of ports along the coast of a country into, e.g., three regions – North, Centre and South (e.g., in LO table). Each of the regions is a stratum and each port belongs to only one of the regions. During sampling, some ports from *each of all three regions* are selected for sampling and estimates for the country built from the individual estimates of each strata.

Similarly to stratified sampling, cluster sampling also considers the sampling units of a sampling level as mutually exclusive groups of units; but contrary to stratified sampling not all these sets are expected to be sampled – only a subset of them is selected for sampling (e.g., randomly). In design based estimation, the estimates from sampled clusters are expanded to the total number of clusters in much the same way as the average or total from a simple random sample are expanded to population size. A simple example of clustering is temporal clustering of days (TE table). Suppose that days in a calendar year are divided into weeks (clusters). The days in each week are mutually exclusive (i.e., each day belongs to one week only). Random sample is used to select a few weeks and in each selected week some days are sampled. If all days in the week are sampled, the design is 1-stage cluster sampling; if only some days are sampled (e.g., 3 randomly selected days in the 7 available) then the design is said to be 2-stage cluster sampling.

Example 1: No clustering ('N')

In Country XYZ the sampling of landings involves a stage of port selection. The ports are organized in three strata: "North" (N = 10 ports), "Center" (N = 20 ports), "South" (N = 50 ports). In a specific time period (e.g., a quarter) 2 ports are sampled from each of these strata.

Locati onId	LOrecor dType	LOstratifi cation	LOclust ering	LOstratu mName	LOcluster Name	LOnumbe rTotal	LOnumberS ampled	LOselection Method	LOinclusi onProb	LOnumberTot alClusters	LOnumberSampl edClusters	LOselectionMet hodCluster	LOinclusionPr obCluster	LOunit Name
1	LO	Y	N	North	---	10	2	SRSWOR	0,2	---	---	---	---	Port L01
2	LO	Y	N	North	---	10	2	SRSWOR	0,2	---	---	---	---	Port L06
3	LO	Y	N	Centre	---	20	2	SRSWOR	0,1	---	---	---	---	Port M07
4	LO	Y	N	Centre	---	20	2	SRSWOR	0,1	---	---	---	---	Port M19
5	LO	Y	N	South	---	50	2	SRSWOR	0,04	---	---	---	---	Port S22
6	LO	Y	N	South	---	50	2	SRSWOR	0,04	---	---	---	---	Port S34

Example 2: 1-stage cluster sampling without stratification ('1C')

In Country Z the sampling of landings involves a stage of port selection. The ports are organized in three geographical clusters: "North" (3 ports), "Center" (4 ports), "South" (5 ports). In a specific time period 2 clusters are selected for sampling and all ports in each cluster are visited. In the example, cluster "Center" and "South" were selected in the sampling and fully inventoried.

Locati onId	LOrecor dType	LOstratifi cation	LOclust ering	LOstratu mName	LOcluster Name	LOnumbe rTotal	LOnumberS ampled	LOselection Method	LOinclusi onProb	LOnumberTot alClusters	LOnumberSampl edClusters	LOselectionMet hodCluster	LOinclusionPr obCluster	LOunit Name
1	LO	N	1C	None	Center	4	4	CENSUS	1	3	2	SRSWOR	0,666666667	Port C1
2	LO	N	1C	None	Center	4	4	CENSUS	1	3	2	SRSWOR	0,666666667	Port C2
3	LO	N	1C	None	Center	4	4	CENSUS	1	3	2	SRSWOR	0,666666667	Port C3
4	LO	N	1C	None	Center	4	4	CENSUS	1	3	2	SRSWOR	0,666666667	Port C4
5	LO	N	1C	None	South	5	5	CENSUS	1	3	2	SRSWOR	0,666666667	Port S1
6	LO	N	1C	None	South	5	5	CENSUS	1	3	2	SRSWOR	0,666666667	Port S2
7	LO	N	1C	None	South	5	5	CENSUS	1	3	2	SRSWOR	0,666666667	Port S3
8	LO	N	1C	None	South	5	5	CENSUS	1	3	2	SRSWOR	0,666666667	Port S4
9	LO	N	1C	None	South	5	5	CENSUS	1	3	2	SRSWOR	0,666666667	Port S5

Example 3: 2-stage cluster sampling without stratification ('2C')

In Country X the sampling of landings involves a stage of port selection. The ports are organized in three geographical clusters: "North" (15 ports), "Center" (20 ports), "South" (10 ports). In a specific time period (e.g., a week) 2 clusters are selected and 2 ports visited in each of them. In the example, cluster "Center" and "South" were selected in the 1st stage of sampling and ports C16, C07, S02, and S14 were selected in the 2nd stage.

LocationId	LOrecordType	LOstratification	LOclustering	LOstratumName	LOclusterName	LOnumberTotal	LOnumberSampled	LOselectionMethod	LOinclusionProb	LOnumberTotalClusters	LOnumberSampledClusters	LOselectionMethodCluster	LOinclusionProbCluster	LOunitName
1	LO	N	2C	None	Center	20	2	SRSWOR	0,1	3	2	SRSWOR	0,666666667	C16
2	LO	N	2C	None	Center	20	2	SRSWOR	0,1	3	2	SRSWOR	0,666666667	C07
3	LO	N	2C	None	South	10	2	SRSWOR	0,2	3	2	SRSWOR	0,666666667	S02
4	LO	N	2C	None	South	10	2	SRSWOR	0,2	3	2	SRSWOR	0,666666667	S14

Example 4: Stratified sampling of 2-stage clusters ('S2C')

In Country Y the sampling of landings involves a stage of port selection. The ports are organized in two geographical strata: "North" and "South". In the stratum "North" there are 8 municipalities (with different numbers of ports each). In stratum "South" there are 12 municipalities (also with different numbers of ports each). In a specific time period (e.g., a week) 2 municipalities are selected in stratum "North" (clusters: NM2 and NM6) and in each municipality 3 ports are randomly selected and visited (NM2 01, NM2 06 and NM2 03; NM6 02, NM6 06, and NM6 03); In the same time period 3 municipalities are selected in stratum "South" (clusters: SM11, SM1 and SM5) and in each of them 2 ports are randomly selected and visited (SM11 03 and SM11 04; SM1 01 and SM1 03; SM5 04 and SM5 05).

LocationId	LOrecordType	LOstratification	LOclustering	LOstratumName	LOclusterName	LOnumberTotal	LOnumberSampled	LOselectionMethod	LOinclusionProb	LOnumberTotalClusters	LOnumberSampledClusters	LOselectionMethodCluster	LOinclusionProbCluster	LOunitName
1	LO	Y	S2C	North	NM2	7	3	SRSWOR	0,4286	8	2	SRSWOR	0,25	NM2 01
2	LO	Y	S2C	North	NM2	7	3	SRSWOR	0,4286	8	2	SRSWOR	0,25	NM2 06
3	LO	Y	S2C	North	NM2	7	3	SRSWOR	0,4286	8	2	SRSWOR	0,25	NM2 03
4	LO	Y	S2C	North	NM6	6	3	SRSWOR	0,5	8	2	SRSWOR	0,25	NM6 02
5	LO	Y	S2C	North	NM6	6	3	SRSWOR	0,5	8	2	SRSWOR	0,25	NM6 06
6	LO	Y	S2C	North	NM6	6	3	SRSWOR	0,5	8	2	SRSWOR	0,25	NM6 03
7	LO	Y	S2C	South	SM11	4	2	SRSWOR	0,5	12	3	SRSWOR	0,25	SM11 03
8	LO	Y	S2C	South	SM11	4	2	SRSWOR	0,5	12	3	SRSWOR	0,25	SM11 04
9	LO	Y	S2C	South	SM1	3	2	SRSWOR	0,6667	12	3	SRSWOR	0,25	SM1 01
10	LO	Y	S2C	South	SM1	3	2	SRSWOR	0,6667	12	3	SRSWOR	0,25	SM1 03
11	LO	Y	S2C	South	SM5	5	2	SRSWOR	0,4	12	3	SRSWOR	0,25	SM5 04
12	LO	Y	S2C	South	SM5	5	2	SRSWOR	0,4	12	3	SRSWOR	0,25	SM5 05