Documentation of the

Regional Database and Estimation System

Data Model

RDBES Data Model doc. v. 1.16

Table of Contents

[Philosophy of the data model 4](#_Toc517418995)

[Key Concepts in RDBES Data Model 5](#_Toc517418996)

[Sampling Data (CS) 5](#_Toc517418997)

[Aggregated Population Data (CE and CL) 6](#_Toc517418998)

[Description of each table 6](#_Toc517418999)

[Upper hierarchy 6](#_Toc517419000)

[Auxiliary tables 7](#_Toc517419001)

[Lower hierarchy tables 7](#_Toc517419002)

[Hierarchy Summary 7](#_Toc517419003)

[The new design variables 10](#_Toc517419004)

[Sampling hierarchies (applies to Design Table; Sample Table) 10](#_Toc517419005)

[Samples (applies to all tables that are part of hierarchies) 10](#_Toc517419006)

[Total units and Sampled units (applies to all tables that are part of hierarchies) 10](#_Toc517419007)

[Stratification (applies to all tables that are part of hierarchies) 10](#_Toc517419008)

[Clustering: One stage and Two stage cluster sampling (applies to all tables that are part of upper hierarchies, except sample) 12](#_Toc517419009)

[Inclusion probabilities (applies to all tables that are part of hierarchies) 13](#_Toc517419010)

[Sample selection methods (applies to all tables that are part of hierarchies) 13](#_Toc517419011)

[Sample Weights 13](#_Toc517419012)

[New features of the RDBES 14](#_Toc517419013)

[True zero values and Non-recording zero values in species sampling 14](#_Toc517419014)

[Non-responses and missing values due to quota sampling 14](#_Toc517419015)

[Other key concepts 15](#_Toc517419016)

[Commercial species 15](#_Toc517419017)

[Percentage catch covered by onboard observations 15](#_Toc517419018)

[Recording of samples collected by fishers or control (new in the RBDES) 15](#_Toc517419019)

[Recording of non-probabilistic sampling (new in the RBDES) 15](#_Toc517419020)

[Annex I: Description of each hierarchy 16](#_Toc517419021)

[Hierarchy 1 At-sea sampling from a vessel list or from a reference fleet 16](#_Toc517419022)

[Hierarchy 2: At-sea sampling from a list of trips 18](#_Toc517419023)

[Hierarchy 3: at-sea sampling where time (e.g., days, weeks) is the primary sampling unit and vessel is the second sampling unit 20](#_Toc517419024)

[Hierarchy 4 On-shore sampling by selecting from location\*time (e.g. harbour-day), then from Fishing Trips, then from Landing Events from those Fishing Trips 22](#_Toc517419025)

[Hierarchy 5 On-shore sampling by selecting from location\*time (e.g. harbour-day) as primary sampling unit, then from Landing Events as the secondary sampling unit. 24](#_Toc517419026)

[Hierarchy 6 On-shore sampling where Fishing Trips are sampled 26](#_Toc517419027)

[Hierarchy 7. On-shore sampling at markets or ports, where location\*time is the primary sampling unit 27](#_Toc517419028)

[Hierarchy 8 Onshore sampling where time (e.g., days, weeks) is the primary sampling unit and vessel is the secondary sampling unit 29](#_Toc517419029)

[Lower Hierarchy A: Length stratified biological samples 31](#_Toc517419030)

[Lower Hierarchy B: Only length frequency data is taken from sample(s)/subsample(s) 32](#_Toc517419031)

[Lower Hierarchy C: All individuals in the sample/subsample are biologically analyzed 33](#_Toc517419032)

[Lower Hierarchy D: No length measurements or biological analyses 34](#_Toc517419033)

[Annex: II RDBES 35](#_Toc517419034)

[Annex III: Frequently asked questions 37](#_Toc517419035)

# 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 v1.0.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.
7. The tables of the **lower hierarchy** can now be populated.

## Aggregated Population Data (CE and CL)

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]](#footnote-2).

# 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  (FO) | This table holds information about the fishing operation – it is very similar to the HH in the 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 we can sample all the landings from a fishing trip due to e.g. landings and sales 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). |
| 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

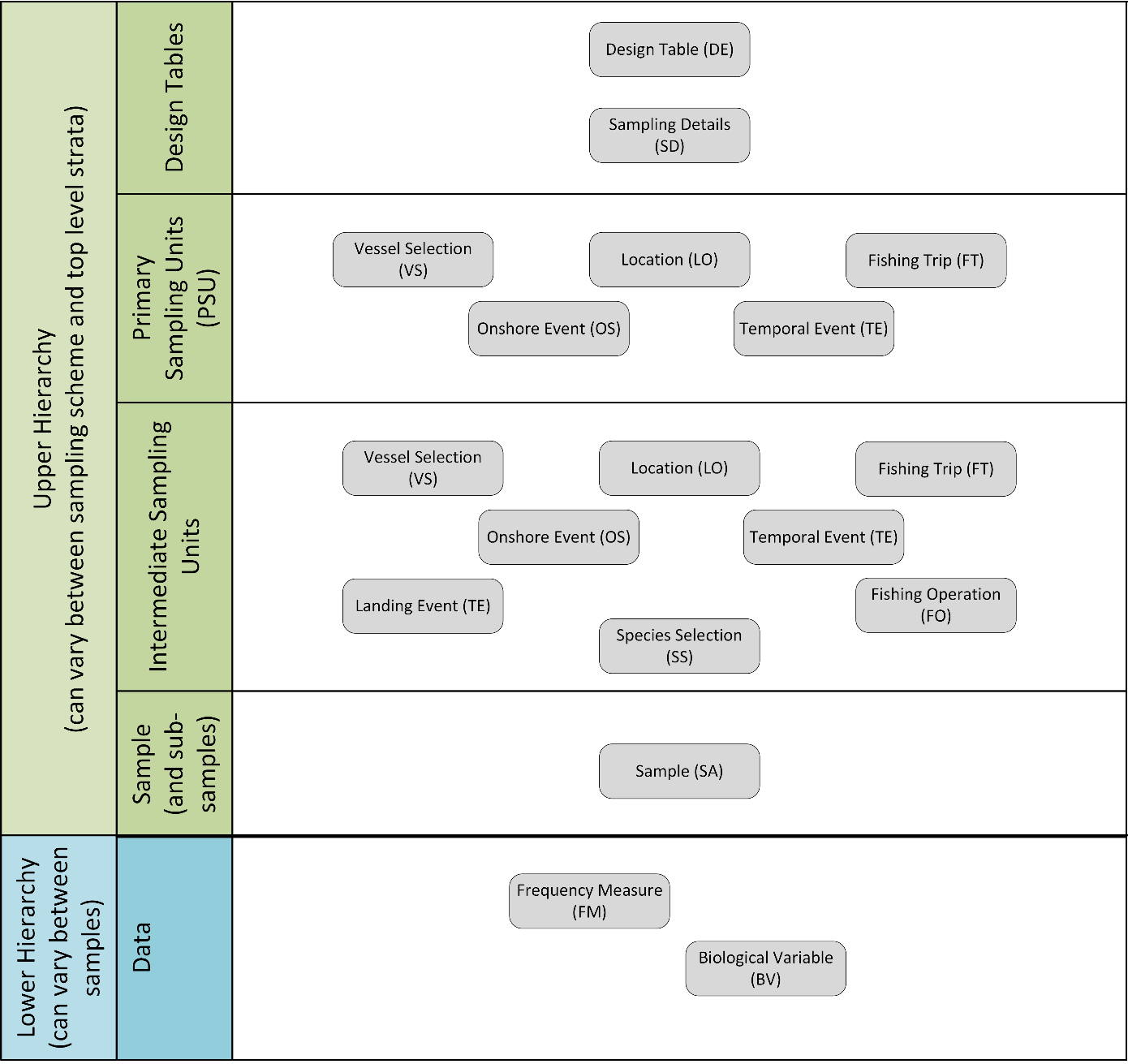
|  |  |  |
| --- | --- | --- |
| **Upper hierarchy auxiliary table** | **Description** | ***be aware*** |
| Sampling details (SD) | This table holds information on the sampling country and institute. | This table is mandatory. |
| Vessel details (VD) | This table holds information about the vessel properties 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 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. | The Single fish number (id) (BVfishId) is what links the different measurements on a single fish together.  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.



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

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Hierarchy 1** | **Hierarchy 2** | **Hierarchy 3** | **Hierarchy 4** | **Hierarchy 5** | **Hierarchy 6** | **Hierarchy 7** | **Hierarchy 8** |
| **Tables in the upper hierarchy** | **Design** | **Design** | **Design** | **Design** | **Design** | **Design** | **Design** | **Design** |
| **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** |
| **Fishing Trip** | **Fishing Operation** | **Vessel Selection** | **Fishing Trip** | **Landing Event** | **Fishing Trip** | **Species Selection** | **Vessel Selection** |
| **Fishing Operation** | **Species Selection** | **Fishing Trip** | **Landing Event** | **Species Selection** | **Species Selection** | **Species List Details** | **Landing Event** |
| **Species Selection** | **Species List Details** | **Fishing Operation** | **Species Selection** | **Species List Details** | **Species List Details** | **Sample** | **Species selection** |
| **Species List Details** | **Sample** | **Species Selection** | **Species List Details** | **Sample** | **Sample** |  | **Species List Details** |
| **Sample** |  | **Species List Details** | **Sample** |  |  |  | **Sample** |
|  |  | **Sample** |  |  |  |  |  |

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** |
| **Design class (WKPICS, 2013)** | **B** | **A** | **Not defined** | **C** | **C** | **C** | **C** | **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 table |  |
| Biological Variable table |  |  |  |

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

## 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 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 total and sampled in all table with table name as prefix e.g. VStotal and VSsampled 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 totalClusters and sampledClusters (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** | **VStotal** | **VSsample** |
| 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** | **VStotal** | **VSsample** |
| 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.
  + 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);

**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** | **VStotal** | **VSsample** |
| 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.
  + 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).

**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** | **VStotal** | **VSsample** |
| 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** | **SAtotal** | **SAsample** |
| 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)

Besides stratification, the RDBES data model can also accommodate one- and two-stage cluster sampling of, e.g., clusters of ports, geographical regions, vessels, etc. At the present moment, cluster sampling is only reported by in a small number of sampling programmes but it is envisioned this number will increase in the coming years. The specification of cluster sampling in the RDBES follows the same line of reasoning as the specification of stratified sampling with the difference being that an additional total and sample value may have to be specified (the total and sampled number of clusters). Detailed examples of stratification (or its absence), 1-stage and 2-stage cluster sampling can be found in “RDBES\_ Stratification\_And\_Clustering Examples.xlsx”

## Inclusion probabilities (applies to all tables that are part of hierarchies)

Each table of an RDBES hierarchy contains a probability column, sampProb. That column should hold the specific inclusion probabilities of each row / sampled unit. In most cases, e.g., when simple random sampling is the selection method of the sample, the column does not need to be filled because the inclusion probabilities can be automatically calculated as n/N where n and N are the sampled and total of each stratum. It is when more complex designs are used, e.g., unequal probability sampling, that values should be specified.

## 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.

## Sample Weights

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

# 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** | **VStotal** | **VSsample** | **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 while SAid 2 and 3 will have child entries.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **SAid** | **SAspeciesCode** | **SAtotal** | **SAsample** | **SAnoSamp-**  **ReasonFM** | **SAreasonNot**  **SampledBV** |
| 1 | Cod | 10 | 2 |  | Non-response – Quota reached |
| 2 | Hake | 10 | 5 | Non-response – sampling Error |  |
| 3 | Haddock | 10 | 5 |  |  |

# Other key concepts

## Commercial species

A frequent need of commercial sampling programmes is that the species composition of boxes from generic commercial designations[[2]](#footnote-3) (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.

## Percentage catch covered by onboard observations[[3]](#footnote-4)

The FO table of the RDBES includes variables where the approximate level of *de facto* observation of the two main onboard activities can be recorded. This is done in two variables – FOpercCoverHaul and FOpercCoverSort – that are used to report the approximate percentage of screening done. The actual significance of these variables varies depending on the fleet. In trawlers, FOpercCoverHaul related to the final hauling of the net and opening of the co-end or activity of the pump; FOpercCoverSort relates to the sorting of the catch by the fishers on the upper or lower deck. In gillnetters and longliners, FOpercCoverHaul relates to the full screening of the haul operation; FOpercCoverSort may (or not) be the same depending on on-board operations.

## 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

## 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 8 upper hierarchy and the 4 lower hierarchies. The first 3 of the 8 upper hierarchies are for at-sea sampling, they are followed by 5 hierarchies for on-shore sampling.

### Hierarchy 1 At-sea sampling from a vessel list or from a reference fleet

Different schemes which all fits the hierarchy 1:

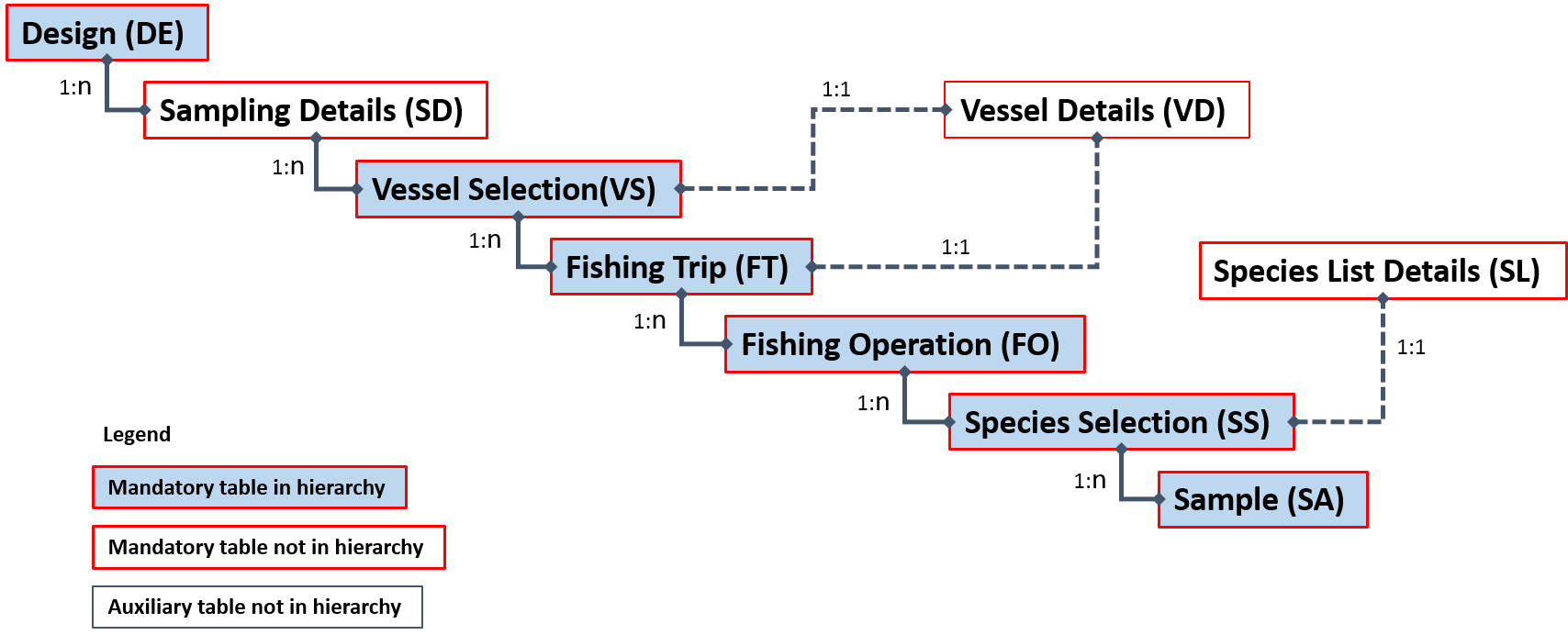
* at-sea sampling with vessel selection from a vessel list, then fishing trip(s) from each vessel. Some or all fishing operations (hauls/sets) are sampled;
* at-sea sampling with vessel selection from a reference fleet, then fishing trip(s) from each vessel. Some or all fishing operations (hauls/sets) 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)

*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)
* Species List Details (SL) (mandatory)



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”.

### Hierarchy 2: At-sea sampling from a list of trips

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

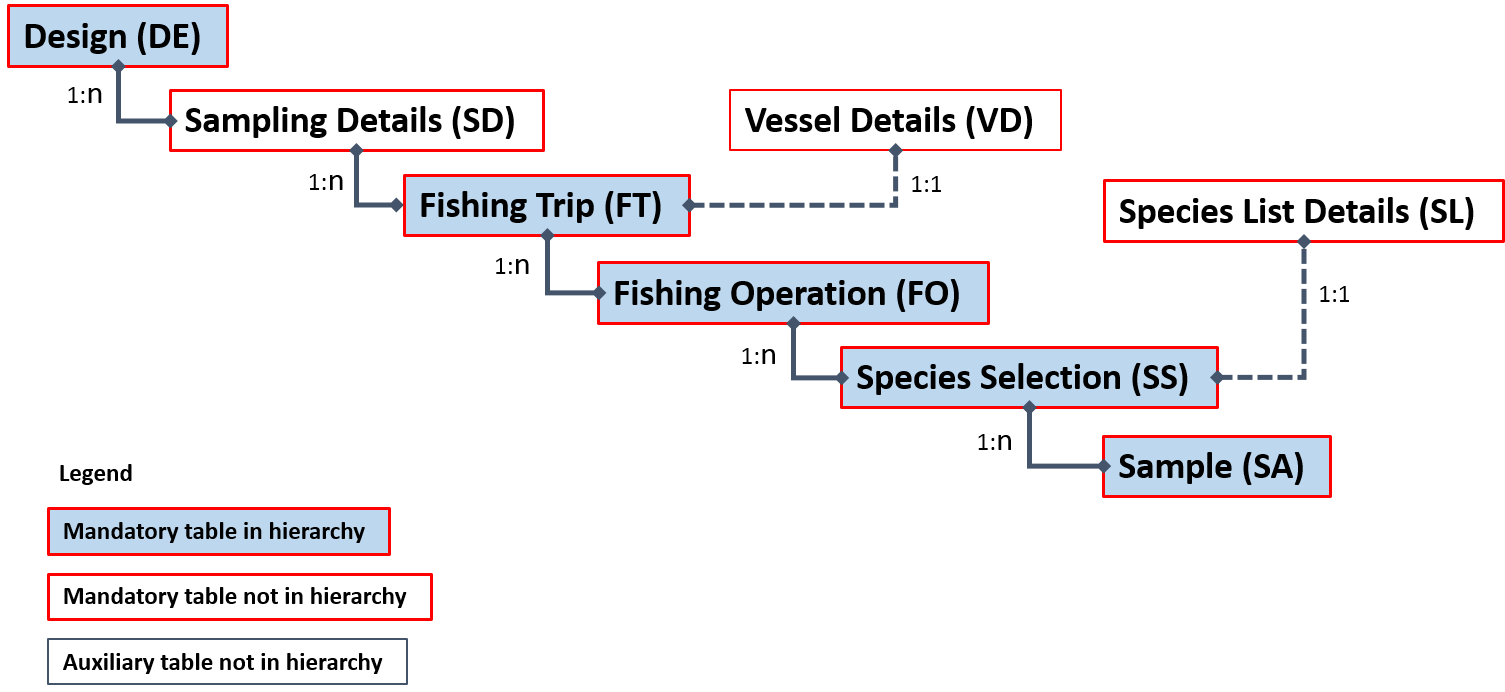
A list of trips to sample from is not always available, but a vessel list is used as a proxy for selecting trips. In reality we often try to squeeze sampling design where we sample from a vessel list into this hierarchy, when it may be more appropriated to use hierarchy 1[[4]](#footnote-5).

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:

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 ‘FoTotal’) – 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: at-sea sampling where time (e.g., days, weeks) is the primary sampling unit and vessel is the second sampling unit

Different schemes fit the hierarchy 3:

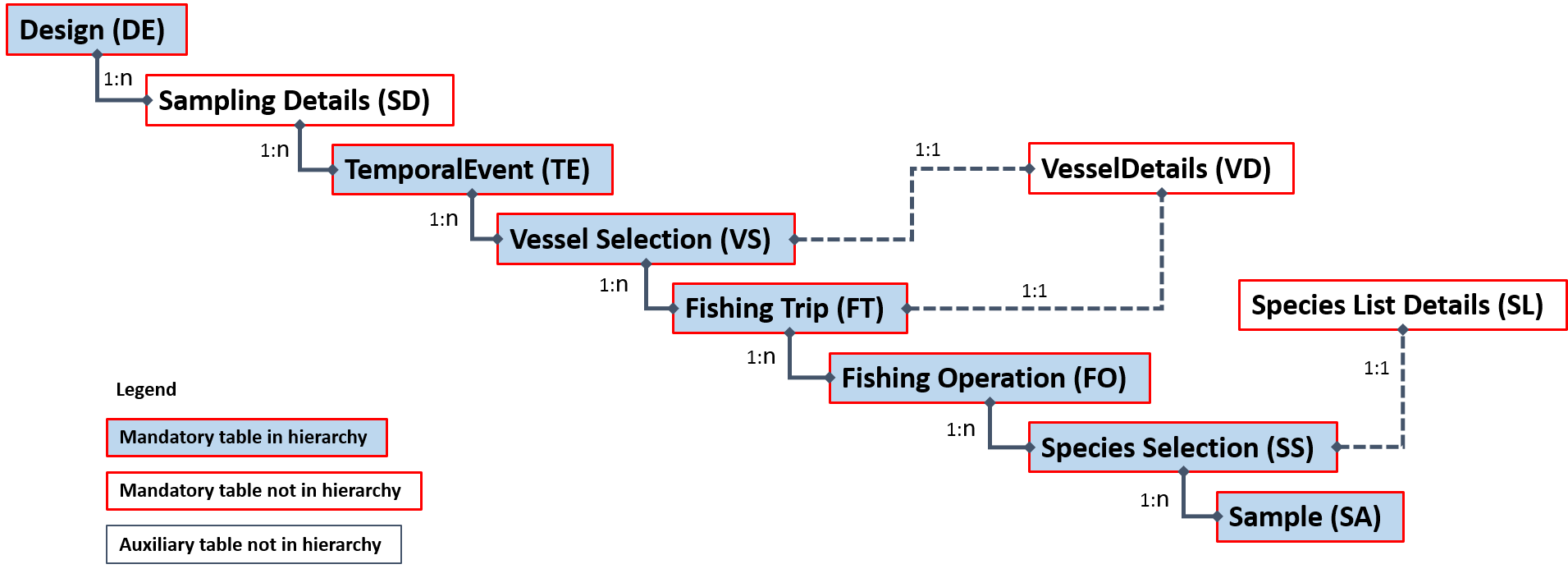
* At-sea sampling where a time period is first selected (e.g., a set of weeks) for sampling to take place. Then, in each time period, vessels in a list are contacted and fishing trips booked. Observers are deployed on fishing trips and sample the fishing operations. In each fishing operation the two catch categories - discards and landings – that are sampled separately and the weights of the species present in each category are determined.
* Variants:
  1. Self-sampling by fishers
  2. Selection of different time units (e.g., day, month)
  3. Sampling all fishing operations or a subset of them
  4. Sampling only discards, taking a catch sample that is sorted later into landings and discards, etc.
  5. Sampling all species or only species from pre-defined list(s)
  6. 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)
* Species List Details (SL) (mandatory)



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 On-shore sampling by selecting from location\*time (e.g. harbour-day), then from Fishing Trips, then from Landing Events from those Fishing Trips

Different schemes which all fits the hierarchy:

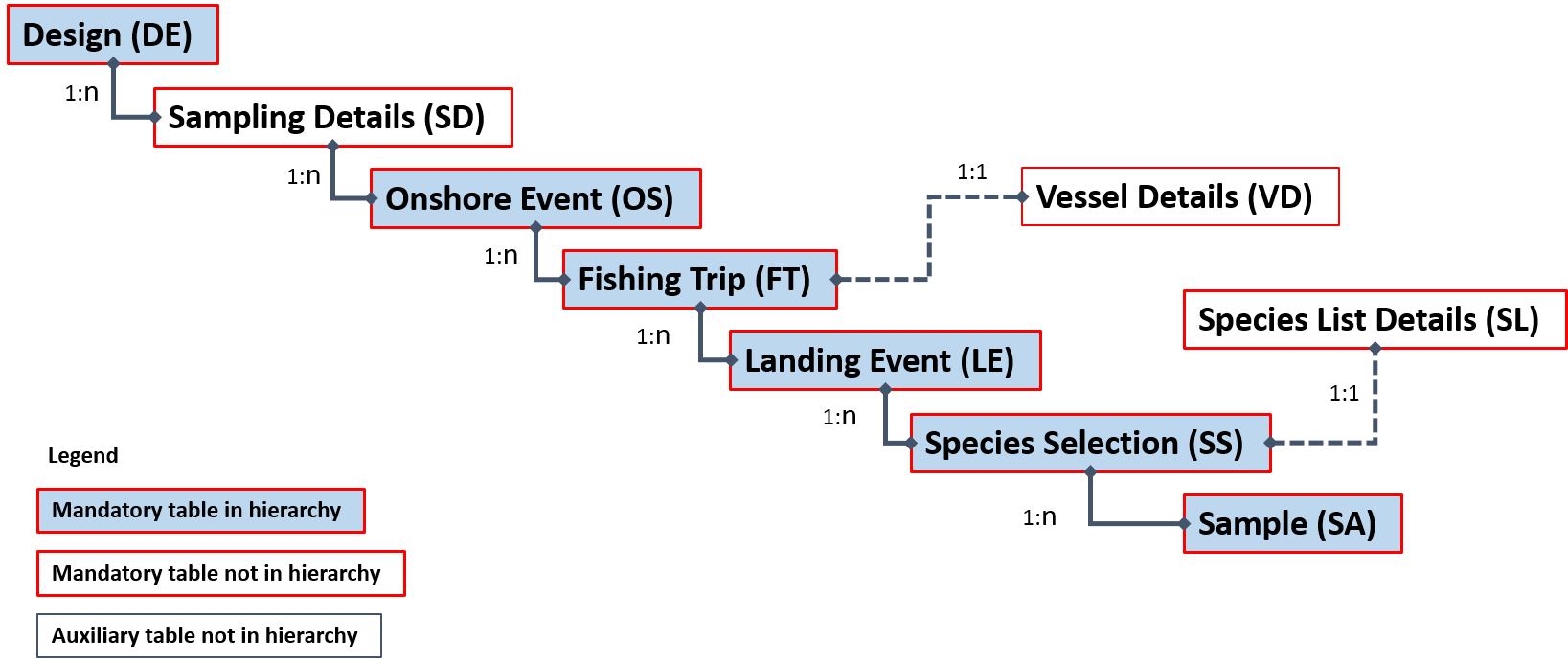
* On-shore sampling at markets or ports. A location and time period (e.g. harbour-day) are chosen and then Fishing Trips that are landing at that location-time period are sampled; detailed information on the fishing trip is available, then Landing Events from those sampled Fishing Trips are sampled.

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

Different schemes which all fits the hierarchy 5:

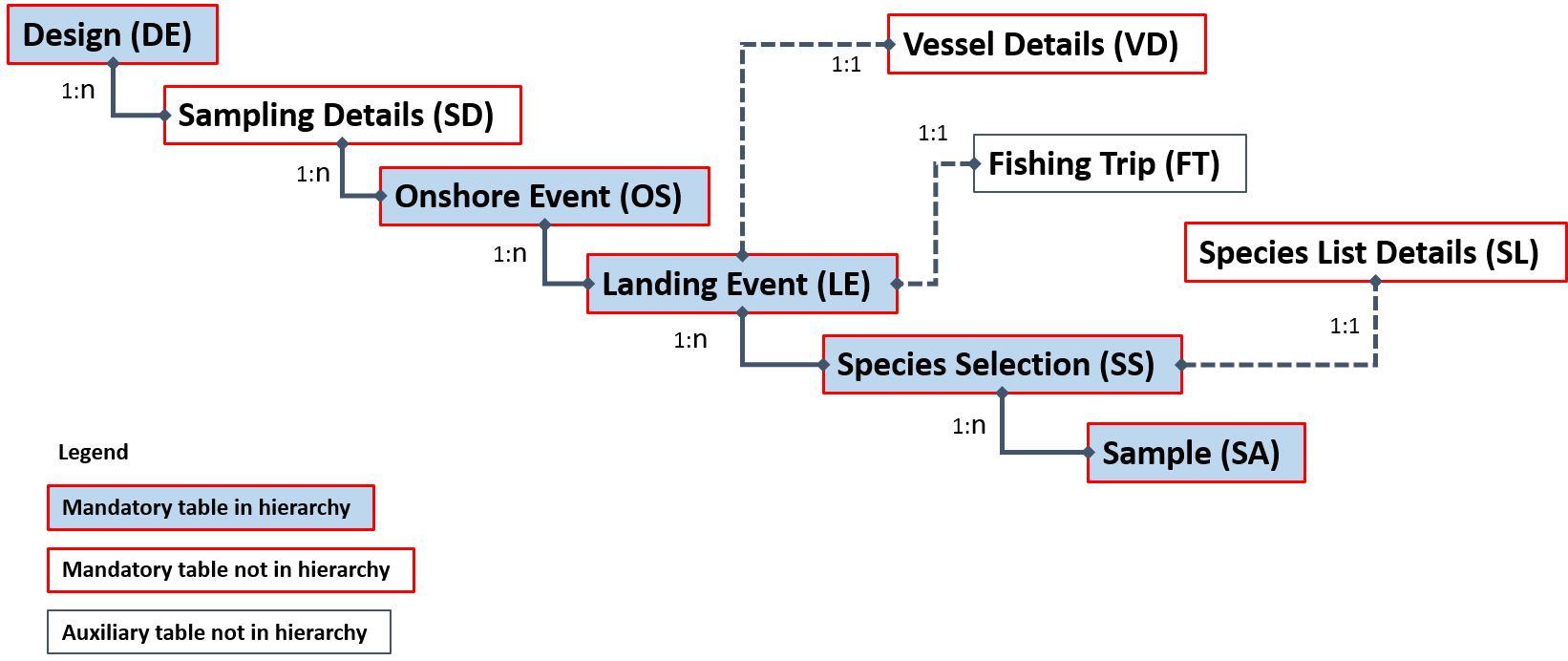
* On-shore sampling from a list of landing sites sampled at random days, samples of all or some landings. Complete fishing trip is known, but not part of the selection procedure.
* On-shore sampling from a list of landing sites sampled at random days, samples of all or some landings. Fishing trip is not known.
* Variants:
  1. Random selection from a list of landing sites.
  2. Random selection from a cluster of landing sites, followed by random selection within clusters.
  3. 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)



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. The landing site is contacted and if landings of target species are expected, it is visited for sampling. All landings are registered stratified by gear, and a species-stratified sample of all catch of the species is taken. From each species sample a subsample of 30 fish is taken for obtaining biological parameters.

Each site-day is registered in the *Onshore Event* table, including those when no landings were expected. Each landing is recorded in the *Landing Event* table. Target species are registered 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 On-shore sampling where Fishing Trips are sampled

Different schemes which all fits the hierarchy 6:

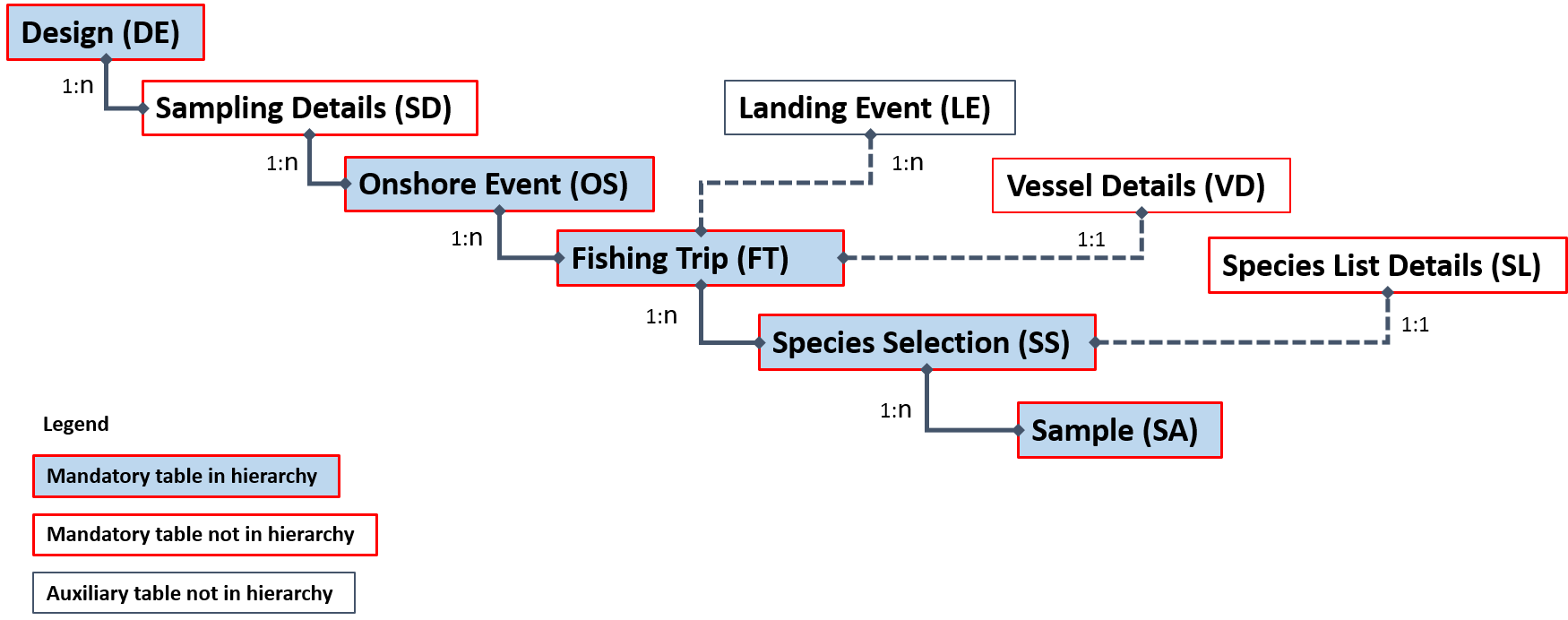
* On-shore sampling at markets or landing ports; known Fishing Trips sampled at random; sale categories and single, concurrent or stratified species selection.
* On-shore sampling from buyers or vendors; known Fishing Trips; sale categories of aggregated landing, and single, concurrent or stratified species selection.

*Upper hierarchy tables:*

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

*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 will 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. On-shore sampling at markets or ports, where location\*time is the primary sampling unit

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:

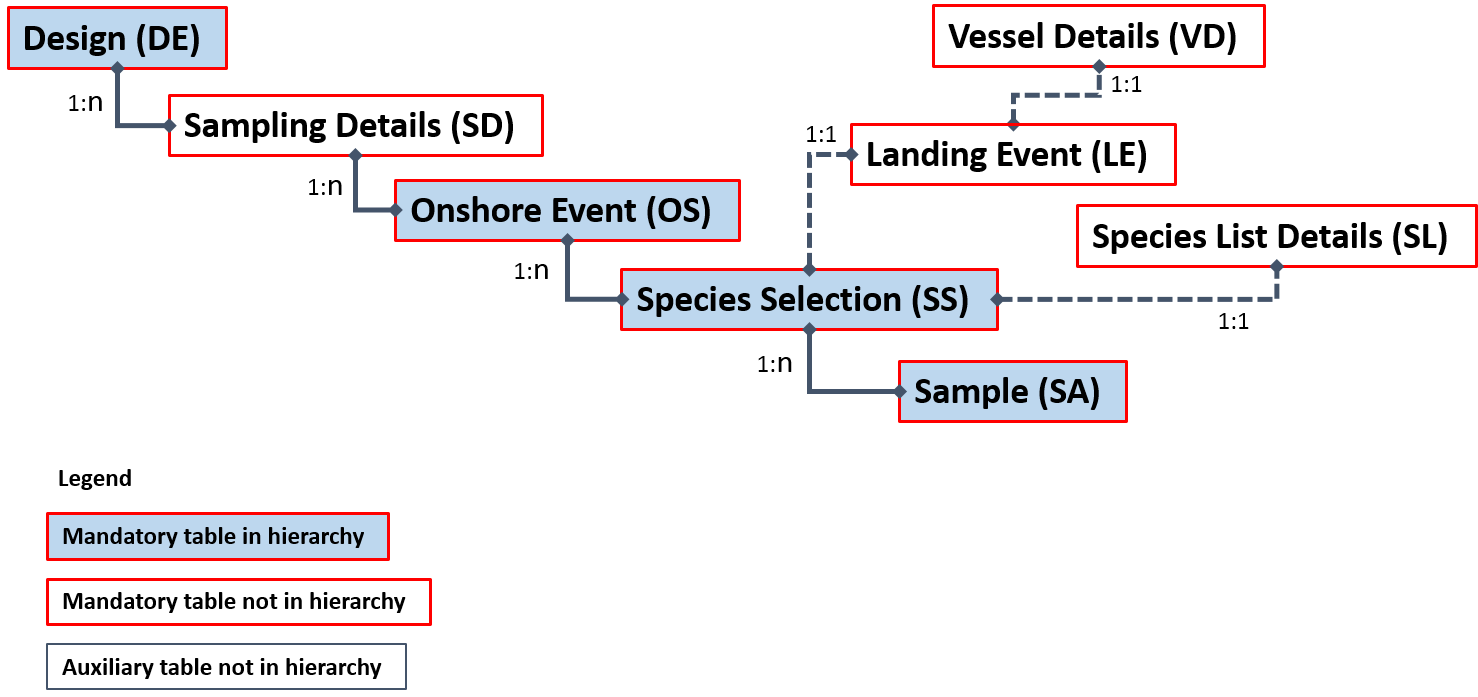
* On-shore sampling at markets or ports. Some or all species present at the market are sampled without considering which fishing trip or landing event they come from. Information on the fishing trip and/or landings events may be collected but is auxiliary.
* On-shore sampling at markets or ports. Information on the fishing trip or landings event is not available. Some or all species of the landing are sampled.
* Variants:
  1. On-shore sampling from buyers or vendors.
  2. 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)



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

Different schemes fit the hierarchy 8:

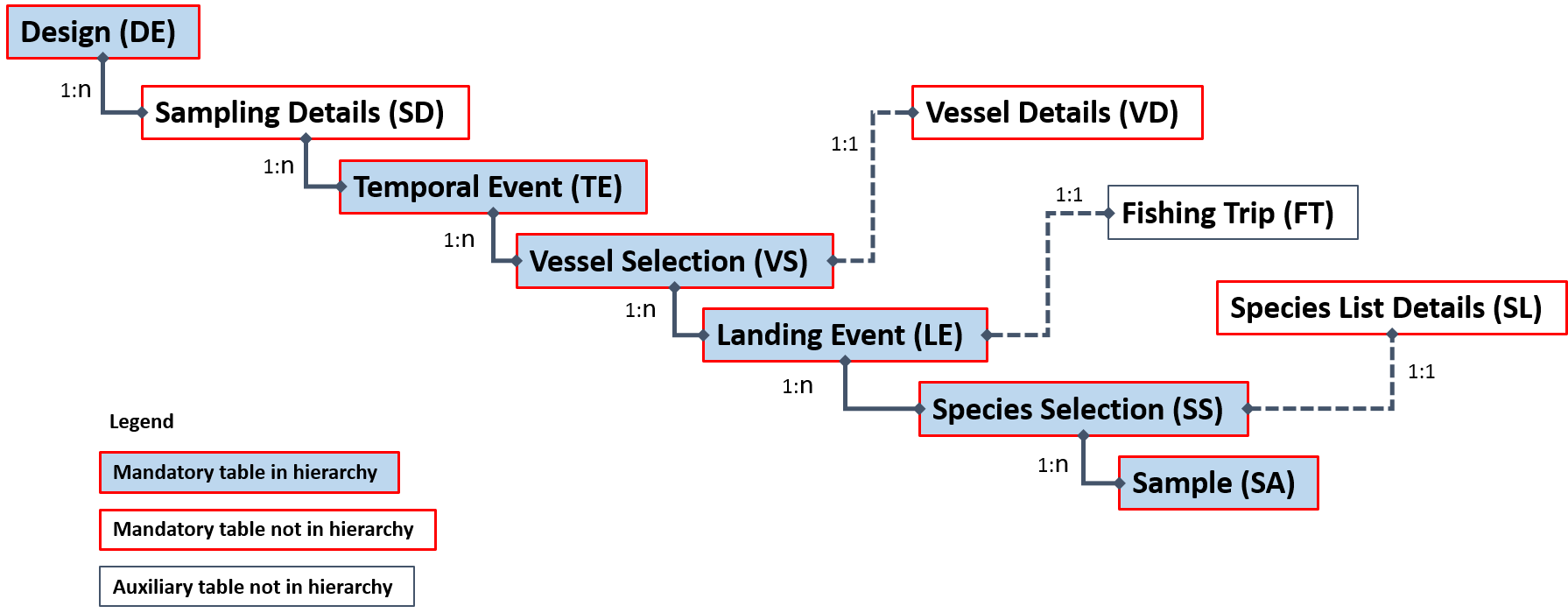
* Onshore sampling of landings where a time period is first selected (e.g., a set of weeks) for sampling to take place. Then, in each time period, a set of vessels is selected for sampling. One (or more) landing events from those vessels are sampled and in each one (or more) species are sampled. Information 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:
  1. Sampling carried out in ports or auctions
  2. Sampling done by observers or subcontractors
  3. Selection of different time units (e.g., day, month)
  4. Sampling all species in the landing or only species from pre-defined list(s)
  5. 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”.

### Lower Hierarchy A: Length stratified biological samples

Different schemes fit the lower hierarchy A:

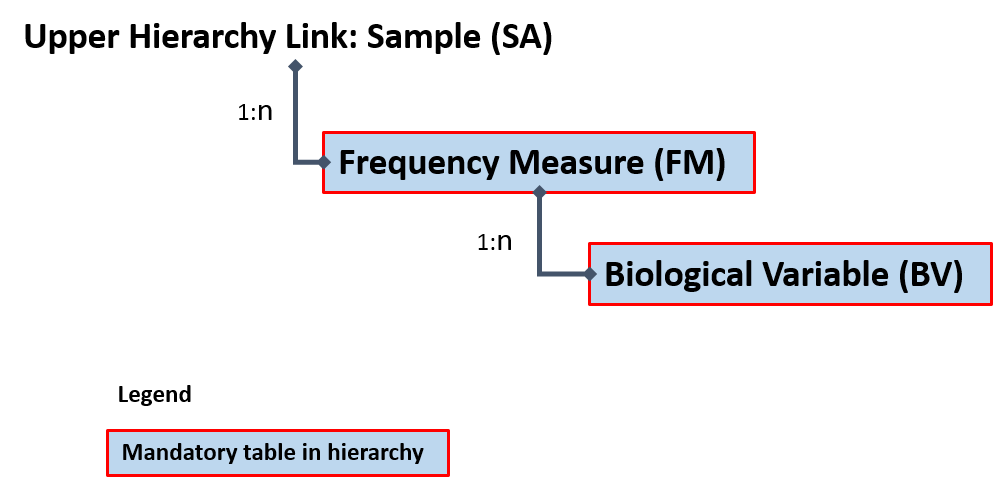
* The sample of a specific species is length measured and stratified by length into 1 cm classes. A fixed number of specimens is biologically sampled from each class (e.g., 2 individuals from each length class). Length, weight, sex and age are determined for each of them.
* Variants:
  1. Different length classes (e.g., 1mm, 0.5cm, 2.5cm, etc.)
  2. Different sample sizes from each length class (e.g., proportional to the number of fish in the length class).
  3. 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:

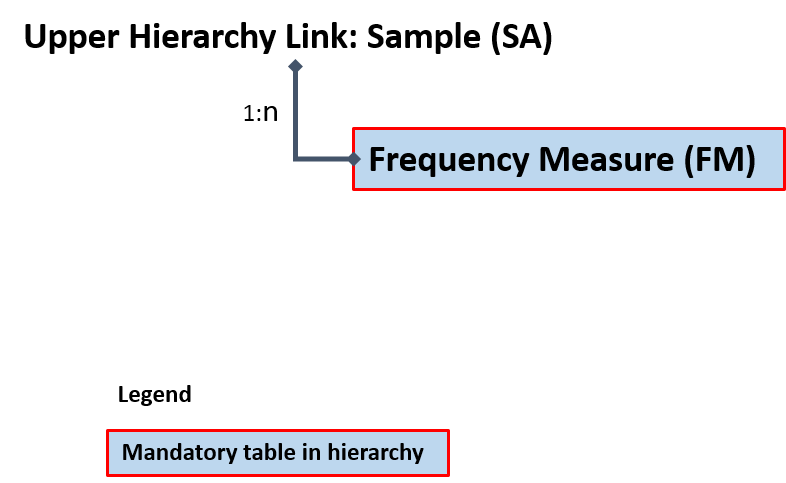
* The sample of a specific species is length measured to the lowest cm. No further biological analyses are carried out on the specimens.
* Variants:
  1. Different length classes (e.g., 1mm, 0.5cm, 2.5cm, etc.)

*Upper hierarchy link:*

* Sample (SA)

*Lower hierarchy tables:*

* Frequency Measure (FM)



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:

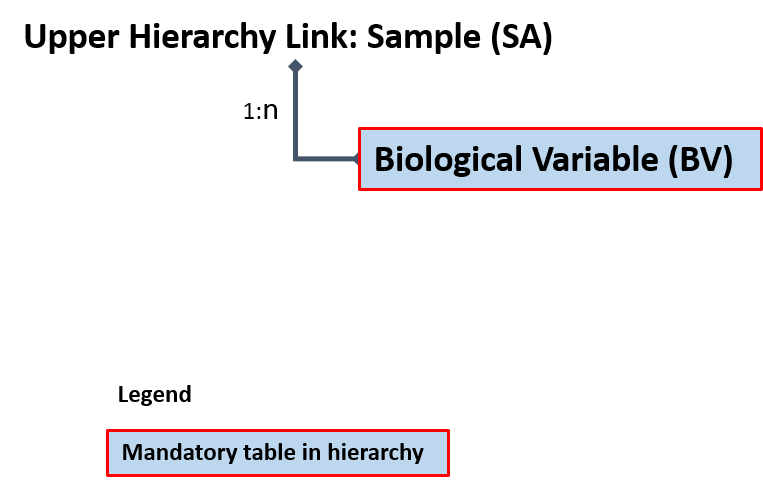
* The sample/subsample of a specific species is entirely biologically analyzed.
* Variants:
  1. Various biological variables collected on each fish sampled (e.g., maturity, stock, age quality, etc.)
  2. Length is not among the variables sampled (e.g., sampling directly for ages)

*Upper hierarchy link:*

* Sample (SA)

*Lower hierarchy tables:*

* Biological Variable (BV)



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 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:

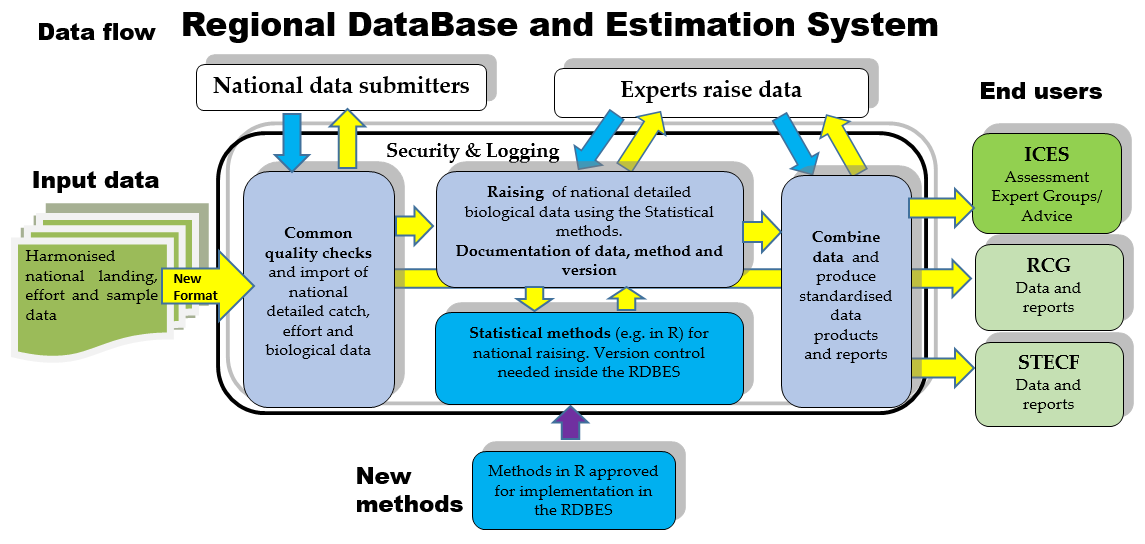
In a sample of species X no fish are measured or biologically analysed.

# 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. When the RDBES 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 RDBES, which is needed for the stock assessment. This will reduce the numbers of data calls, which will benefit the countries. The countries are also interested ensuring the estimated data for the ICES stock assessment is correct, that is done by the RDBES only are using approved estimation methods, which cannot be instantly manipulated. All the processes will be logged and the processes will be transparent.

The approach should be to call the statistical methods written in R, which have been encapsulated into the RDB using version control. The encapsulation of the methods into the RDB is important, because this will ensure the methods are approved and it is not possible to modify the encapsulated methods written in R inside the RDB. Having the raising methods defined in R would make the raising more transparent and easier for the experts to update, if needed. It should be possible to download both the data and the methods from the RDB, so the experts easily can mimic the raising in the RDB and further develop the methods, see the diagram below.



*Diagram of the data flow and processes including the estimation using R.*

When a group of experts have developed a new statistical raising method or updated and existing method, the group should approach the WGCATCH or a subgroup thereof, which should have the task to test and approve raising methods. When the method have been approved, the method will be encapsulated into the RDB, using the RDB’s version control of methods. The method can now be used to raise uploaded data, but the method cannot be manipulated/edited, and the raised data will be transparent and fully documented, regarding data and what method and version that was used.

# Annex III: Frequently asked questions

**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 ‘FoTotal’ and ‘FoSampled’ 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 ‘FoTotal’ and ‘FoSampled’ 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: ‘SAtotal‘, ‘SAsampled’, ‘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: ‘SAtotal‘, ‘SAsampled’, ‘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:** what is the different between fishing trip and sampling event in onshore sampling schemes?  
**Answer:** When observers visit a port in a specific day (i.e., an onshore event) several situations may take place, e.g., observers may take direct contact with the trips and vessels collecting data on landings (and possibly on other fractions) [hierarchy XX]; or observers ... [KIRSTEN TO COMPLETE]

**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** | **VStotal** | **VSsampled** | **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 |

1. http://www.ices.dk/marine-data/data-portals/Pages/RDB-FishFrame.aspx [↑](#footnote-ref-2)
2. hereby termed commercial species for sake of simplicity [↑](#footnote-ref-3)
3. This section is particularly relevant for the recording incidental by-catch and will likely suffer reformulation after comments from WKPETSAMP and WGBYC. [↑](#footnote-ref-4)
4. “**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) [↑](#footnote-ref-5)