EU MAP: Data Collection Framework Methodology for Fisheries & Aquaculture

Socio-Economic Data

March 2023

BIM Ireland's Seafood Development Agency

Contents

Fisheries

Data Sources

Economic and social data for fleet segments are sources from the following data sets:

- Sales notes data for landing income for vessels under 10m.
- Logbook declarations data for effort and landing income for vessels over 10m.
- Voluntary questionnaire information returned by vessel owners targeted in the annual economic survey, the National Seafood Survey (NSS) for all economic and social variables.
- Face-to-face/phone interviews with vessel owners to clarify any issues arising with economic and social variables from questionnaire.
- Mandatory economic and social questionnaire information returned by vessel owners applying for EU/National grant aid.
- Data from vessel owners from a national sentinel vessel programme (to collect both transversal and non-transversal economic and social data from vessels in the small-scale fisheries where log-book declarations are not mandatory).

Data Collection Methods

Data collected from the EU fleet register, Logbooks and Sales notes is treated as census data for capacity, landings, and effort data reporting.

Given the constraints imposed by the voluntary nature of the current data collection regime, the data collection scheme for economic variables from all metiers segments is a non-probability sample survey (NSS) based on a probability sample survey design.

In 2010, a Statutory Instrument[1] (S.I. 132 of 2010) was introduced by the Member State (MS) requiring all fishery sector operators to collect and maintain economic data as listed in Annex XII of the Commission Decision. However, there is no enforcement of this national legislation.

Sampling Frame

The population shall be all active and inactive vessels registered in the Union Fishing Fleet Register as defined in Commission Implementing Regulation (EU) 2017/218 on the Union fishing fleet register on 31 December of the reporting year and vessels that do not appear on the Register at that date but have fished at least one day during the reporting year.

[1] http://www.irishstatutebook.ie/eli/2010/si/132/made/en/print

The data sources for the national implementation for the fleet target population are:

- EU Fleet register.
- EU Log-book data.
- Sales notes data.

The target population is the "commercial fishing fleet" as recorded in the EU Fleet Register on the last day of the reference year and any other vessels that were active and recorded on logbook, sales notes, or national fleet register data sets.

Fleet Segmentation: The segmentation of the fleet, follows the guidelines in Table 8 of Commission Delegated Decision (EU) 2021/1167 and is used to stratify the collection of all, non-transversal, economic parameters.

The following data sources will be used to segment the fleet:

- EU Fleet Register on the 31st of December for the reference year,
- EU log-book activity records for vessels active in the reference year (>10 meters)
- Sentinel Vessel Programme (SVP) effort data,
- Recorded fishing activity from previous economic surveys.

Individual vessels are assigned to fleet segments by overall length (LOA) class and the main fishing method engaged in by the vessel, in the previous calendar year. In cases where there is a risk of natural persons and/or legal entities being identified clustering may be applied to report economic variables to ensure statistical confidentiality. Such a clustering scheme shall be consistent over time.

The source of information used to distinguish the sampling frame from the target population, will be based on EU logbook data as follows:

- Active vessels: For vessels greater than 10 meters in overall length, only those with at least one entry in the EU logbook, in the reference year, will be deemed active. This analysis will take place once the log-book data are available for the reference year, which is approximately 3 months after the end of the calendar year (March n-1).

- For vessels less than 10 meters in overall length, an estimate of inactivity will be conducted each year using all available sources, including: previous survey responses, the national inshore SVP, sales notes data and the fleet register.

Required sampling intensities have been estimated using statistical analysis of the previous year's survey data. The analysis determines required sample size n, based on the mean of a finite population, to achieve a given level of precision (e.g., a coefficient of variation (CV) of 25% on the sample mean).

Applying the function, we can see that for extremely low CV, all vessels need to be sampled and that the required sample number increases with the standard deviation of the segment. However, due to the finite population function you can never sample more than the full population (census). Some segments have a planned sample rate of 0% as the number of active vessels in the segment are extremely low (n=<5). These segments have been presented in their entirety in Table 3A (Sample Rate 'N') but will most likely be clustered with similar segments which have higher number of vessels for data submission.

Sampling

The sampling strategy is of a stratified, random design. In the reference year, economic data will be collected from 15% of the total fleet. Data will be collected from 33% of vessels in fleet segments > 12 metres in length (LOA), and 10% of vessels in fleet segments < 12 metres in length (LOA).

Although no specific precision levels are prescribed (in the Regulation) for economic variables, the sampling programme adopted by the MS is constructed to achieve a precision of 25% at a 95% confidence level, in line with Level 2 (set out in Chapter two) of Commission Decision (2010/93/EU)[2]. Further stratification of the fleet in order to refine economic analyses, is theoretically possible but practically impossible to achieve. Such stratification and improvement of the sampling framework will, however, be addressed by EU MAP staff practised and dedicated in such work, in anticipation of legislative change.

Determination of sample size for each fleet segment

Required sampling intensities have been estimated by statistical analysis of historical survey data from the last four years of surveys [1]. The method uses a random sampling formula (1) to determine how many samples (n) are required to obtain a pre-determined precision level on a parameter where σ2 is the sample variance, d (2) is the required precision of the estimate in the same units as the mean, and t is the t-value. A more detailed explanation can be found in Appendix 1.

[2]COMMISSION DECISION of 18 December 2009. Adopting a multiannual Community programme for the collection, management, and use of data in the fisheries sector for the period 2011-2013 (notified under document C(2009) 10121) (2010/93/EU)

[1] Minto, C. (1998) Precision level estimates: robust to small sample sizes of finite populations for the Data Collection Regulations.

$$n = \frac{t^2\sigma^2}{d^2} \quad (1) \qquad d = t_{\alpha/2,n-1}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \quad (2)$$

The MS conducts an annual survey in which all active vessels are requested to submit economic, employment and operational details for their previous year's activity.

Under the terms and conditions of EMFAF co-funded grant schemes contained in the National Seafood Development Operational Programme and implemented in the reference year, all vessel owners in receipt of grant-aid will be required to comply with the annual National Seafood Survey.

Following advice and input from the industry, the timing of the annual survey is now scheduled to coincide with the final date for submission of statutory tax returns for the previous financial year (i.e., the survey opens in October and runs until the following February). This is intended to encourage the active participation of fishermen and their accountant's in providing the relevant data at a single and convenient time of the year.

Appropriate application of sampling theory will direct the evolution of the sampling framework. Although the voluntary nature of the annual survey prevents the practical development of the sampling frame, the development of such innovation will represent a core function of the DCF staff group. It is hoped that as more surveys are returned through grant aid support, which will require the applicant to allow BIM access to their accounts for a five-year period, will allow us to create a targeted random sample of these vessels.

Raising and estimating procedures

Recognising the implications and influences imposed by the voluntary nature of the annual survey on the probability sample survey design standard appropriate raising techniques will be used, to derive final estimates for each variable collected. This methodology was reviewed in 2018, which resulted in a report to assess and improve the raising estimations. Various methods of raising are possible, and this report set to establish a theoretical and empirical basis for the decision as to how best to raise sampled economic data to the fleet level.

Raising consists of multiplying a summary statistic from a sample of respondents by a measure of the total quantity for the population. This could be an average raised by the total number or amount of

effort or capacity comprising the sampled vessels. The mean squared error (MSE) encapsulates the bias and variance of an estimator. The MSE was used as the basis for comparing raising performance. We first derive theoretical expectations on which raising method would work best when there is or is not a relationship with fishing effort. Raising methods were then tested on the real data via re-sampling and appraisal of the ability of various raising methods to recover the true sum. A suite of specifically developed visualisation code assists in appraising the distribution of the data, in particular with identifying outlying values that can overly influence the raised sum.

Economic variables independent of effort

From the theoretical analyses there were two major conclusions for raising sample data:

1. MSE of the raised average is derived in Appendix (A.1) and given by:

$$\mathrm{MSE}(\hat{y}_{\mathrm{avg}}) = N^2 \frac{\sigma_x^2}{n},$$

where σ^2_x is the marginal variance of the economic variable.

2. MSE of the effort raised sum is derived and given by:

$$\mathrm{MSE}(\hat{y}_{\mathrm{eff}}) = \frac{\sigma_E^2}{\mu_E^2}\left(\frac{1}{n} - \frac{1}{N}\right)\left(\mu_x^2 N^2 + N^2 \frac{\sigma_x^2}{n}\right) + N^2 \frac{\sigma_x^2}{n},$$

where μ_E and σ^2_E are the mean and marginal variance of the effort variable, respectively; and μ_x is the mean of the economic variable.

Unless $n = N$, the MSE of the effort raised sum will be greater than the average raised sum. This is because the variance of the estimator is inflated by the inclusion of an unrelated variable. Thus, the average raised sum is a better estimator when there is no relationship between the economic variable and effort.

Economic variable unrelated to effort

When the economic variable is independent of effort, a raised value based on the average is a better estimator than raising by effort.

Economic variable is proportional to effort.

1. MSE of the raised average is given by:

$$\mathrm{MSE}(\hat{y}_{\mathrm{avg}}) = a^2 N^2 \frac{\sigma_E^2}{n} + N^2 \frac{\sigma_x^2}{n},$$

where a is the proportionality constant between the economic variable and effort; here σ^2_x is the variance of the residuals around the proportional relationship between the economic variable and effort.

2. MSE of the effort raised sum is given by:

$$\mathrm{MSE}(\hat{y}_{\mathrm{eff}}) = a^2 N^2 \frac{\sigma_E^2}{N} + N^2 \frac{\sigma_x^2}{n} \left(\frac{\sigma_E^2}{\mu_E^2} \left(\frac{1}{n} - \frac{1}{N} \right) + 1 \right).$$

Where $n = N$ the two approaches equate (as expected). To simplify the relationship, we derive (in Appendix A.2) an inequality where the MSE of the raised average is a higher (worse estimator) than the effort raised when there is a proportional relationship with effort. This is given by:

$$\frac{\sigma_x}{a\mu_{\tilde{e}}} < \sqrt{n}$$

The left hand-side encapsulates the ratio of how variable the relationship with effort is. If it is very variable and the relationship with effort is weak, a poorer estimator is expected by raising based on effort but if the relationship is strong (small residual variability and large proportionality constant) raising based on effort is a better approach.

When the economic variable is proportional to effort, the improvement in the performance of raising based on effort depends on the strength and variability of the relationship with effort. Where the relationship is weak and variable and the sample size small, the estimator based on raised effort may be poorer; whereas for a strong relationship with small residual variance, raising based on effort will result in an improved estimate.

A plot of the MSE under both approaches helps to illustrate the overall theoretical finding that when there is a strong relationship with effort, raising based on effort provides a better estimator; conversely when there is no relationship with effort, raising based on the average is better (Figure 1).

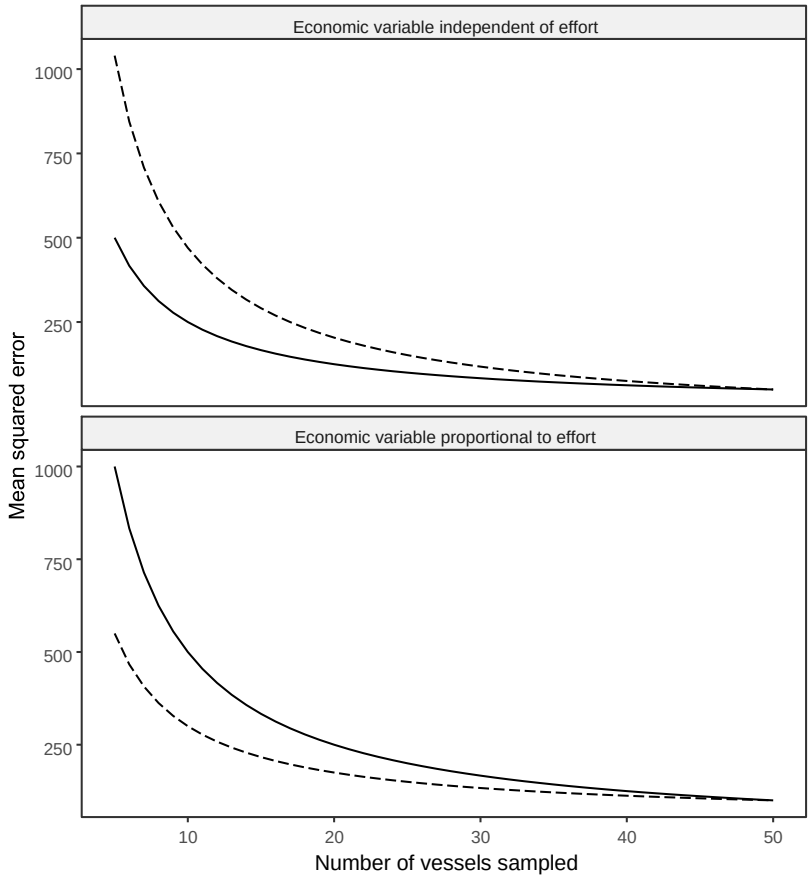

Figure 1: Illustration of MSE when there is or is not a relationship with effort. *x*-axis is the number of vessels sampled *n* out of a total fleet of $N = 50$; *y*-axis is the MSE (average squared difference between the true and estimated value) based on the derived equations for when the sum is raised based on the average or by effort. Values used for illustration were: $\mu_X = 1$, $\sigma_x = 1$, $\mu_E = 1$ (top panel), $\sigma_E = 1$, and $a = 1$ (bottom panel).

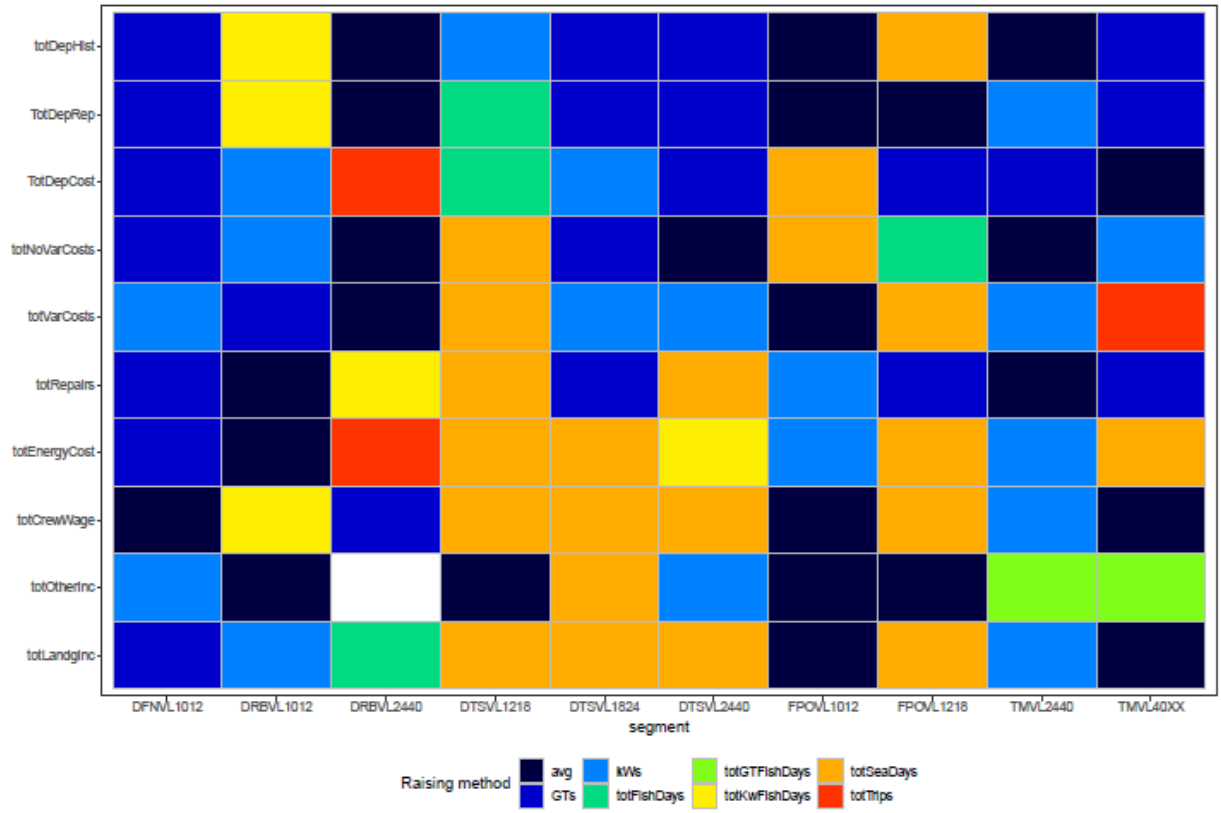

Figure 2. Best performing method (highest mean rank) by variable and segment

Capital Costs

In accordance with Appendix VI of Commission Decision (2008/199/EC), a variation of the Perpetual Inventory Method (PIM) will be applied to estimate capital value and costs for each of the fleet segments. In line with the EU MAP guidance document alternative methods of capital costs valuations can be based on company surveys. These alternative methods may be used if the derived estimates reflect the actual definition of net capital stock (depreciated replacement value of the vessel including on-board equipment with a useful lifetime of more than one year).

Capacity indicators and capital value will be estimated for all vessels on the register, regardless of their activity. The following sources will be used to estimate the input parameters to the PIM model:

- Questions on fixed assets, investments, and depreciation from the annual economic survey,
- EU fleet register,
- EU log-book data,
- Sentinel vessel programme.

Ireland is planning to conduct more work on this to adapt current capital valuations.

A harmonised FTE will be estimated for each of the fleet segments. For vessels >10 meters in length, operational data from log-book submissions will be used in the estimation of fishing time on a trip-by-trip basis. In addition, there are several questions on the annual economic survey forms that deal specifically with hours worked and the nature of the engagement (full-time, part-time, casual). Questions regarding gender breakdown and age profiles, education and nationality will behave been added to the annual survey. A full review of this methodology is planned in 2023/24.

Data Quality

The sampling intensity is based on an analysis of the variance of historic, operational data, as these have proved to be the most uniform, with the aim of achieving a precision of 25% at a 95% confidence level. Variances within fixed costs have proved much higher than expected and, as such, quality will be measured at a coverage rate, commensurate with the target precision for the non-operational parameters. The Member State conducts regular reviews of its statistical methodologies to ensure the highest standards in data quality

While traditionally the NSS for fisheries was conducted by a postal survey, in 2021 the MS launched a new an online portal to make the process of providing data easier and more secure. Data for the 2022-2027 Work Plan will be submitted electronically via an online survey portal to a centralised database, with pre-validation necessary before the accountants can successfully submit the data also via an online survey portal. A secondary validation process will be performed on the data once received, and any erroneous data will be queried directly with the vessel owners or their accountants, by survey personnel. Similarly, any erroneous data supplied by vessel owners, contracted under the sentinel vessel programme, will be queried and rectified by survey personnel, as and when it arises, or at the exit interview stage of the programme.

Although error associated with bias and variability will effectively be introduced if observed returns do not match those expected, these descriptors will be reported where possible and with appropriate caveats.

The issue of consistency of data coming from different data sources is recognised as being of significant importance. The introduction of bias in this area, is under continual assessment and is currently being addressed by restricting acceptance of data to a small number of official data streams (i.e., data items consistent with fields in annual company returns (provided via accountants), EU logbook data and Sales notes data).

Social Data

Following the pilot study in 2018 for the collection of socio-economic data these variables were incorporated into the annual economic survey. In general, there was good response to the questions on by age, and employment status. However, employment by education level, was difficult to collect but has also been added to the annual surveys.

Data Quality

The sampling intensity is based on an analysis of the variance of historic, operational data, as these have proved to be the most uniform, with the aim of achieving a precision of 25% at a 95% confidence level, in line with Level 2 of Commission Decision (2010/93/EU)). Variances within fixed costs have proved much higher than expected and, as such, quality will be measured at a coverage rate, commensurate with the target precision for the non-operational parameters. The Member State will seek further statistical advice on this aspect of the National Programme.

A general improvement in the quality of data received from vessel owners has been realised, consistent with the implementation of the following operational measures;

- Ensuring that a qualified accountant signs off on the financial aspects of the survey;
- Timing the survey to the coincide with the end-of-year submission of tax returns to ensure all figures are already checked for the Revenue Commissioner's Office;
- Requesting full, end-of-year, accountant's reports.

Although error associated with bias and variability will effectively be introduced if observed returns do not match those expected, these descriptors will be reported where possible and with appropriate caveats.

The issue of consistency of data coming from different data sources is recognized as being significant. The introduction of bias in this area, is under continual assessment and is currently being addressed by restricting acceptance of data to a small number of official data streams (i.e. data items consistent with fields in annual company returns (provided via accountants), EU logbook data and Sales notes data).

Aquaculture

Background

Before EU-MAP/DCF requirements, aquaculture production and employment data had been gathered by census for national purposes, from the late 1990s. The national format in terms of data segmentation, was found to be compatible with DCF, as well as Eurostat and FAO requirements. The continuing census is collected, as before, on a voluntary basis and therefore relies heavily on goodwill. The survey conducted historically by post, now utilises a combination of interactive and static online and phone surveys.

Census returns historically are over 80% of the total population by entity number and provided basic production and employment data. Non-returns thus far have received no sanction from appropriate authorities and estimates for these were obtained indirectly from local aquaculture area officers, from other agencies or from estimates made using 'Indicator' companies and/or historical records.

Data collection strategy since the introduction of DCF/ EU MAP

Most Irish operators are small and access to accounts and costs data is expensive, difficult to extract (e.g., bottom mussel seed costs) and time consuming. Therefore, the data collection strategy for DCF economic data was devised to minimise the extra burden being placed on clients and consists of a single annual survey program to cater for the needs of national programs, as well as to satisfy the data requirements of DCF and other end-users. A common pool of raw data is thus gathered to supply all requirements.

The Annual survey of year N commences in early January of year N+1.

The survey consists of two components; a census survey by questionnaire of producers for all required input-output, volumes, incomes and costs data, plus a non-random, sample survey of company accounts, available online to obtain balance sheet data. Abridged accounts are surveyed from November of year N+1 as these become available for basic balance sheet data.

The census survey is launched in early January of the following production year by email and website activation. A six- week deadline is given for return of the census though in practice, it takes over two months to receive most of returns, with a trickle of data continuing to arrive into May.

A phone survey commences close to the original deadline and continues for up to a month afterwards while late returns will be accommodated as far as possible. Publications of the production data occurs

in March (Business of Seafood) and July (Business of Aquaculture) or earlier if requested. Data is then prepared for FAO and Eurostat from June.

Survey of abridged accounts online occurs from the end of November while the survey forms for the following year are prepared/designed. Database maintenance is ongoing and design review/repair occurs before the next survey launch.

Frame Population and Surveys

The Population: All active, legally producing businesses are part of the population surveyed annually. There is no threshold applied.

The statistical unit: The lowest level of data disaggregation is the Production Unit (PU) which is the subunit of a business relating to a specific Species/culture enterprise, in one location, from which a distinct turnover and associated employment figure can be generated. Ideally this refers to a distinct sea-site within a bay or an onshore or inland facility, into which, a distinct Spp. stock is input and upon which an aquaculture practice is imposed until the stock is harvested and sold. A business therefore, farming two species or one species by two different methods is considered to have a minimum of two production units. The majority of Irish operations are one-PU entities.

A collection of licenced production sites within one bay can be considered as belonging to one production unit, provided only one Business is operating them to culture a distinct stock, sites are homogenous and annual reporting of the sites activities in this way is consistent.

'Enterprise' is a term applied more loosely than Production unit (PU). An enterprise can be a one or multiple PU business activity. It can be a sole trading or company business. Generally, it is a term applied to a particular location and activity of a business, but the Production Unit is meant to be specifically defined to serve, as far as possible, as a distinct business activity, at one location and is the statistical unit upon which the survey census questionnaire is designed.

An active business is one where one or more persons are employed or self-employed, applying or intending to apply an aquaculture practice upon livestock for the purpose of sale for financial gain, either within the surveyed year or subsequent to that year. Some active businesses do not produce a turnover every year.

Aquaculture business

An aquaculture business is a primary producer; one that inputs into a licenced site or facility, an aquatic species livestock, then grows and / or imposes an aquaculture practice upon that stock, to a point where the stock is harvested and sold, either live or dead from the farm gate.

- The 'Farm gate' sale is a vital part of the producer designation.
- Producers are striving to value add to their original bulk product. Some primary preparation of raw product, such as filleting or small-quantity packaging is evolving on some sites. In these cases, the enterprise is still considered to be a primary producer, rather than a processor.
- 'Processing businesses' are those dedicated to processing activity. They have purpose-built and equipped facilities wherein they apply fundamental changes to the raw material they purchase to create their final products. In several cases, a parent company will have a producing company, a processing company, and a trading company.
- In the case of self-contained production – Processing facilities, every effort is made to dis-aggregate the production from the processing components of the data and then to populate the DCF segment templates appropriately. In the case of 'financial variables, this can be challenging.

DCF Data Collection Strategy

- Variable data collected by census: These are production; scale, input volumes and costs, output volumes and values, Production unit numbers by employment categories and employment number by category. The new environmental and socio-employment variables are also collected by census survey.

- Variable data collected by online sampling: These are balance sheet financial variables such as 'Depreciation', 'Assets' Liability' 'net investment' 'financial costs' and 'Wages & Salary' that are readily available by this source despite the abridged nature of these micro-business accounts.

Certain variables are obtained by more than one source (primary and secondary sources) each source can be used to aid in validation of data from another.

Variables derived:

These are: 'Imputed value of unpaid labour',' FTE' and for finfish 'Mortality.'

Method for calculating 'FTE':

- Full Time: >30 hrs/week or > 40 weeks a year
- Part Time: 10-30 hrs /week or 13-39 weeks * 40hrs a year
- Casual: > 10 hrs /week or < 13 weeks * 40hrs a year

'Imputed value of unpaid labour':

This is estimated for each sampled business, then for each segment sample, then estimated for the national segment.

Minimum expected value for 'wages and salaries' for the segment is calculated by:

Segment FTE * national minimum wage for the sampled year

Actual 'wages and salaries' value for the segment is obtained by survey.

The two values are compared:

If 'Actual value' >= minimum expected then no unpaid labour value

If 'Actual value' < minimum expected then the difference = 'imputed value of unpaid labour'

Timeframe:

- Jan N +1: launch of census questionnaire-based survey by notification and supply of questionnaires by email and on the BIM website.
- March N+1: Census data collected used for publication of annual 'Business of Seafood.
- April N+1: Census survey, active phase. is wound up and data is used for publication of Annual 'Business of Aquaculture' report. Note that raw data arriving later is accepted and annual estimates are updated, causing some discrepancy between reports by different institutions. The argument for late data use is that the latest data available will be used to answer any legitimate query. The importance of prompt data supply however is emphasised to clients.
- June N+1: Census data is used and formatted to supply requirements of the OECD.
- Methods reviewed and checked for deviations from the DCF national plan. National plan tables are reviewed and updated.
- July N+1: Census data is used and formatted to supply requirements of the FAO.

- August N+1: Census data used and formatted for first draft of data to supply Eurostat requirements.
- November N+1: On-line survey of abridged accounts begins. Accounts are downloaded and financial and any other usable variable data is extracted.
- Nov-December N+1: final upload of Eurostat data
- January-February N+2: on-line survey of year N abridged accounts is wound up.
- March N+2: Sample data of year N is raised to national level. First draft of DCF data for year N is ready for upload if necessary. Data is rechecked through for errors one month before upload call deadline.

Census data assessment and estimation

Generally, there are 80% returns from the census. Estimates of non-returnee data is made from one or more of the following methods depending on the aquaculture sector and individual history.

Sector trend and five- year historical average: For certain segments such as the fresh rope mussel sector, unit value is consistent within a bay, therefore the unit value trend among compliant neighbours can be applied to the five-year average unit value of the non-compliant PU.

Structural data: Direct returns are the preferred data source; however, the return must be plausible for the capacity of the PU making the return. Consultation with the local area officer, if not directly with the Producer is required if reported production does not match perceived capacity.

Assessment of area officer or other in-house data: Whether because of an unrealistic return or no return, the input of an area officer familiar with the enterprise can eliminate false returns or provide good estimates gleaned while engaged on other agency business with the producer. In-house data gathered from recent specific projects can give a good indication of a non -returning production unit scale and output.

Datasets of other agencies: The aggregated data from other datasets can fill survey data voids but also can be used as validation tools. Mussel seed data from one sister agency for example is preferred over that gathered directly by the DCF survey as that data is gathered at the point of seed extraction, rather than from a later office estimation. The directly gathered oyster seed data is preferred over that of another sister agency as the latter is of the intended seed transfer, rather than the actual, measured by the DCF survey. The other dataset however is a welcome validation aid to the DCF survey data.

Last year's data return or estimate: At the start of the annual survey, the default setting estimate is that production and employment is the same for the year about to be surveyed as the year just surveyed. Subsequently that assumption is normally eliminated as other estimation methods are used or direct return data is obtained. If pushed on occasion for provisional estimations ahead of scheduled reporting, the default assumption is used though only if other estimation options are unavailable.

Non-random sampling of financial variables available online:

A 33% sample of the populations abridged accounts is gathered from on-line sources, as far as possible, across all DCF segments, for financial and some costs and employment data. The sample taken here is gathered at no extra burden to clients and allows the annual sampling of indicator companies whose variable data can be used to track trends directly. This is useful for those segments with segment-dominating businesses.

Data Quality

Sample data is obtained at business level, census data is obtained at Production unit level.

Most businesses are single production unit entities though some comprise several production units of one or more aquaculture practices. The census is designed to receive data at the level of the production unit. Each filled census form is intended to describe the economic activity of one PU. Each Sampled abridged account however is a report to company level only. The data therein is then farther disaggregated to PU level at the data entry point.

4.3 DCF/ EU MAP estimates; contrast with the estimates for National, Eurostat and other institution reports.

Essentially these occur for two reasons, disparity in the questions asked and disparity in when the answers are requested by each institution.

There is significant disparity in the segmentation between DCF and Eurostat datasets and strenuous efforts have been underway to harmonise these for the two different Regulations (EU) 2017/1004 previously (EC) 199/2008) and (EC) 762/2008. Essentially the two Regulations however are not asking exactly the same questions. For example, EU MAP does not distinguish between consumer ready and

products destined for further aquaculture practice whereas Eurostat and FAO do. Additional differences between Eurostat and FAO questions include the means in which juvenile production is recorded.

There is significant disparity in when the data is requested by the different institutions. Theoretically the earliest deadlines receive the cruder estimates, based on less available raw data. Eurostat requires data for year N, from December of year N+1, DCF from approximately September of year N+2 and the collecting agency BIM requires the data from as early as February of year N+1 for the publication of the annual 'Business of Seafood' report. Disparity among all these reports is inevitable as estimation updates will continue to evolve as late data comes in.

Data validation

Data validation is achieved in the following ways:

Comparison of received data with known production capacity:

A producer may supply data that does not match their capacity.

Comparison of received data with the estimations of local area officers.

For Census data, regional estimations are made known to the appropriate area officers who compare these to their own estimations, based on their on-the-ground observations of production level at the Production Units involved. In the case of significant disparity, the producer(s) concerned are contacted and failing this, the datasets of other agencies are compared if available, as well as an appraisal of the known PU production capacity before finally settling on the appropriate estimation of that PUs production data.

Comparison of variable data obtained by different collection strategies

Turnover and basic employment data for a given business, obtained by census, is compared to data obtained from on-line accounts. Only accounts using the calendar year can be used to get an accurate comparison of data, otherwise time lag between data generation will produce discrepancies.

'Wages and Salaries' and 'Depreciation' data, obtained from the census questionnaire can be compared with that obtained from the online Accounts survey. Again, only calendar year accounts can be used for realistic comparison.

Raw materials volume estimates obtained from census can be compared with the data sets of other agencies. In the case where the census data is deemed inferior to that of the other dataset, the latter data is used for DCF reporting. This is the case for example with mussel seed input data.

Estimation; Raising of sampling data

The census generated national turnover data is principally used in raising the value of financial and costs data, gathered by sample, to a national estimate. Critical to the use of turnover data in this way is the assumption that there is a direct relationship between the values of these variables for a given business and the value of that business' turnover.

National turnover is used to assign a percentage value to the turnover reported or estimated for each production unit. This proportion is then used to assign the same proportional value to costs and financial data obtained for each Production Unit sampled. For each such variable therefore, each sample return value also has a value proportional to the national value for that variable. The sum of the sample value for each variable is divided by the sum of assigned proportional values of each PU responding, then multiplied by 100 to raise the sample value to a population estimate for that variable.

(Variable Sample sum / sum of % s of national turnover of each responding PU turnover) * 100)

A disadvantage of this methodology is that, for very small samples obtained for a given variable, the figure calculated is likely to be an over-estimate. Estimation procedures are periodically reviewed.

Errors; detection and correction

Potential for error lies throughout the process but there are particular points where errors are most likely to occur.

Data supplied: A producer may supply data that does not match their capacity. The discrepancy may be due to a typing error or to an aquaculture husbandry practice unknown to the surveyor. Knowledge of the production unit capacity will draw attention to the anomaly if one exists. Contact with the producer should make the farming practice allowing the generation of such data understood or clarify if there was a typing error.

Data clarity: Data-entry by the respondent can be mis-understood, which can be resolved through contact with the producer. Questionnaire formatting left free to encourage a response is susceptible

to ambiguous data supply. Formatting of answering cells are therefore reviewed to see if restricting their formats are feasible to reduce entry choice and therefore increase clarity of data supplied.

Data upload: Data can be entered wrongly by the surveyed and the surveyor when uploading if compelled to upload data manually. Errors may produce an anomaly in the summary tables designed to amplify them which is investigated in the raw data entry and then to the return received and corrected, if necessary, after clarification from the producer.

Database manipulation: The Raw data file is susceptible to damage during query processing. This potential damage is reduced using pivot table querying where possible, rather than manipulating the raw data spreadsheet itself. In the latter case, copies of the raw data page are made and manipulated accordingly.