



Statistics Committee
14th Meeting
11 April 2018
Mexico City, Mexico

**Progress report on establishing a
comprehensive statistical database**

Background

1. The International Coffee Council, at its 120th Session, approved the Five-Year Action Plan for the International Coffee Organization (ICO) ([ICC-120-11](#)). The aim of the Action Plan was to set out the overall direction of the ICO for the next five years and to provide the context for the development of annual programmes of activities in order to prioritize and allocate relevant resources.
2. The Programme of Activities for the Organization for coffee year 2017/18 ([ICC-120-12](#)) was approved by the Council during its 120th Session and covers the activities to be undertaken in order to achieve the Organization's Strategic Goals. The first Strategic Goal outlined in the Action Plan is 'Delivering world-class data, analysis and information to the industry and policy-makers'.
3. One of the activities in the Programme for coffee year 2017/18 is to 'Establish a comprehensive statistical database, containing information relevant to Members and third parties (paying subscribers, wider public)'.
4. In order to establish a comprehensive statistical database, several activities are being undertaken. These include:
 - (a) Conducting a demand survey with users of data (Members & third parties) to review and update the portfolio of variables to be collected by the ICO.
 - (b) Assessing the quality of the current ICO database.
 - (c) Drafting a proposal for a new data management system, including an upgrade of the database, interfaces and processes with the aim of increasing efficiency of data input, quality, handling, storage and dissemination.
5. This report describes the steps taken by the ICO Secretariat, in close consultation with Members, with regard to the three activities. After consideration of the initial results further steps will be taken in the latter half of coffee year 2017/18.

Action

The Statistics Committee is requested to take note of this report.

**PROGRESS REPORT ON ESTABLISHING A
COMPREHENSIVE STATISTICAL DATABASE
MARCH 2018**

Part 1: Demand Survey

1. A demand survey was carried out from 29 November 2017 through 15 December 2017 in order to review the portfolio of variables to be collected by the ICO, as well as ensuring the statistical data currently provided was in line with the needs and priorities of its users.

Methodology

2. An online questionnaire, entitled 'ICO Statistical Questionnaire' was launched on 29 November 2017 inviting Members and third parties related to the ICO's activities to participate in the survey. In addition, the questionnaire was also made available through the ICO's social media and website. The questionnaire was accessible to Members and third parties in the four official languages of the ICO. The online survey was closed on 15 December 2017 and the Secretariat received 312 valid responses, 66 of which were responses from individuals representing Member countries. The questionnaire focused on assessing the importance of 42 different variables that are either regularly collected by the ICO or frequently requested. The questionnaire also allowed respondents to list up to three additional variables, not already included in the questionnaire, that they considered important.

3. The 42 variables included in the questionnaire fall under the following broad categories: production, exports, imports, consumption, prices, stocks/inventories, and socioeconomics. Respondents were asked to rate the level of importance for each variable on a 5-point scale ranging from 'not important at all' to 'extremely important'.

Results

4. All respondents found that all 42 variables were at least moderately important. The difference in importance for each variable generally did not differ much between the public (including Members) and the private sector. This confirms that the data that the ICO currently collects is important for both Members and third parties and should continue to be included in the database. The responses also identified the need for additional variables, particularly socioeconomic indicators, in order to become a more comprehensive database.

5. A ranking of the importance of variables was derived from the variation of the importance levels chosen for each variable. A score was assigned to each variable, weighted by the frequency of responses for each level of importance. For example, a variable for which

every response was 'extremely important' received a score of 5, while a variable for which every response was 'not at all important' received a score of 1. This methodology was applied to each variable, and the resulting scores range from 3.56 to 4.52. Of the 10 variables with the highest score, two ('Cost of Production' and 'Coffee farming profitability') are not currently collected regularly by the ICO. See Annex II for the full list of variables and their rank.

6. There were 111 valid responses in all four languages, in which Members and third parties suggested three additional variables that they considered important to include in the database. The additional variables listed in response to this question fell into a number of broad categories relating to agronomics, climate change, projects, coffee production financing, components of production costs, value addition and information on coffee traders.

7. In summary, the analysis of the response to the ICO statistical questionnaire shows:

- Most of the variables that the ICO currently collects are important to the users confirming the relevance of the current database.
- Additional variables, such as cost of production and employment, that are currently not collected in the ICO database, are being considered as important.

8. Now that the demand survey has been completed, the next stage in this process will look at possibilities for including some of the additional variables into a new, more comprehensive database.

Part 2: Data quality

9. A data quality analysis of the current ICO statistical database was implemented as a first step towards a proposal for a new data management system. This system would increase the efficiency of data input, handling, storage and dissemination as well as improving the overall quality of the ICO's coffee statistics. Through a systematic assessment, the data quality analysis provides a better understanding of the characteristics and quality of ICO data, as well as identifies the primary areas of concern with the data. Further, the results of the demand survey indicate that many of the variables that the ICO currently collects are important to users, and the data quality analysis will help to ensure that this information is of high quality for users.

10. The ICO Administrative Budget for the financial year 2017/18 includes provision for the hiring of an external consultant to perform the data quality analysis and to assist in drafting the proposal for a new data management system. The firm Data-to-Value that was selected as achieving the highest score against the selected criteria among three offers, commenced its assignment in late January.

Methodology

11. The approach of Data-to-Value is multi-step, beginning with building a comprehensive data catalogue and a model of the database. The data asset catalogue and model provide a summary of all of the information in the ICO database, describe individual data items in readily understood terms, and demonstrate how different data are related to each other. This information enables users to quickly find data they need and will be necessary for implementing a new database management system.

12. The next step entailed assessing the quality of the statistical data based on the following parameters: accuracy, completeness, consistency, conformity, integrity and timeliness. During the quality assessment, the data was methodically analysed to discover patterns, identify duplicates and missing values, and confirm whether the data matched expectations. Data patterns that differed from expectations were brought to the attention of ICO Secretariat to determine whether the pattern arose from issues with the quality of the data or were expected values. This step went through several iterations, starting with the largest issues (i.e. that is data that was highly interconnected with other data records or covered a large portion of data records).

Results

13. The analysis looked at both the technical components of the system (i.e. hardware, SQL server) as well as the collection and handling of data for the current database. The current statistical data management system is running on Windows server 2003 and SQL Server 2000, which is more than 15 years old. Based on the initial findings of this analysis, the current design of the data management system demonstrates that the database is generally well-organized and robust. However, one of the major issues of the current system is that much of the database hardware is no longer supported and will need to be upgraded in the near future. Upgrading the hardware will also require upgrading the software to make it compatible. Implementing a new database will not only improve the security of the database, but will also enable the ICO to take advantage of advances in technology that have happened in the last fifteen years to improve the experience of ICO data users.

14. The analysis of the statistical data uncovered unusual data patterns that identified a few items that might be of lower quality. Many of these patterns result from missing data or the inability of the current system to handle the more complex data that the Secretariat receives. The current data management system was initially designed for a limited purpose and fewer variables. Upgrading the system will allow for more flexibility in meeting the changing needs of Members and other users as well as improving the efficiency of collecting and processing raw data, which will result in an improvement in the overall quality of the data.

Summary

15. The key elements of the analysis are:
- (a) The current system is obsolete both as to hardware and software.
 - (b) No support service is available for parts of the current system.
 - (c) Data integrity is at risk.
 - (d) There is a need to add new variables to better respond to the demand of Members and the world coffee community.
16. The data quality analysis of the ICO database will be completed in mid-April. The final results of this analysis, steps taken to remedy issues that were identified, and a proposal for a new data management system will be presented to Members at the September 2018 Council Session.

Conclusion

17. The process of establishing a comprehensive statistical database was initiated by reaching relevant stakeholders through the demand survey and performing a data quality analysis as described above.
18. As a result of the demand survey and data quality analysis, the following was found:
- Input provided by Members as well as third parties interested in the global coffee sector through the statistical questionnaire demonstrated that ICO data is a fundamental source of information and also helped to identify new variables that the ICO should consider adding to its database.
 - The data quality analysis carried out by the consultancy identified unusual data patterns that will be addressed by the Secretariat in order to improve the overall quality of the database over the remainder of this coffee year and in the following coffee year. The identified level of importance for the different variables from the demand survey will assist the Secretariat in prioritizing which data quality issues to focus on when improving the quality of the data¹.
 - The data quality analysis found that the database needs to be upgraded both to improve efficiency in data handling as well as the security. The analysis also generated the knowledge and documentation needed for upgrading the current data management system.

¹ The Secretariat is also working with the Statistics Roundtable to identify and address data quality issues.

19. At the 122nd Session of the International Coffee Council, the Secretariat will present to Members a detailed report and a proposal for their consideration:

- A final list of all variables to be collected by the ICO based on an analysis by the Secretariat that will consider both the findings of the demand survey and the accessibility of data for these new variables.
- A proposal for a new data management system that takes into account not only the results of the data quality assessment of the current database, but also the inclusion of new variables, based on the outcomes of the demand survey. The proposal will also consider ways to facilitate the provision of data from Members as well as assessing and reporting on the quality of the data provided.
- The new database will reflect an improvement in the overall quality of the current ICO data and increase the efficiency of data input, handling, storage and dissemination.

STATISTICAL INFORMATION QUESTIONNAIRE - MEMBERS

Thank you for taking the time to complete this questionnaire on which statistical variables are most important to you. Your responses will help the International Coffee Organization to improve the quality of its statistical information and ensure that it best addresses the needs and priorities of stakeholders in the coffee sector. The deadline for responses is 8 December 2017.

1. Please choose the option which best describes your organization or affiliation: *

- Public sector (government agency, international organization, etc.)
- Academia/Research
- Private Sector
- Other

If you selected "Other" above, please specify here:

2. Do you currently subscribe to ICO statistical reports (Monthly Trade Statistics, Quarterly Statistical Bulletin or Annual Trade Statistics)? *

- Yes No

3. How familiar are you with ICO statistical data? *

- Not at all familiar
- Slightly familiar
- Moderately familiar
- Very familiar
- Extremely familiar

4. The ICO provides data on global, regional, and national levels. What level or levels are you interested in? (Tick all that apply). *

- a. Global
- b. Regional
- c. National

5. How important are the following items relating to coffee production? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Production by volume	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Production by type (Arabica or Robusta)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Production by processing method (dry or wet)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Quality of coffee produced	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Varities/Cultivars produced (e.g. Gesha, Java, Bourbon)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Volume of certified coffee produced (e.g. Fair Trade, etc.)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Total coffee production area	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Total number of coffee trees	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Number of coffee trees in production	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input checked="" type="radio"/> 4	<input type="radio"/> 5
Age of coffee trees					

6. How important are the following items relating to coffee exports? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Exports by volume	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Exports by value	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Exports by type (Arabica, Robusta)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Exports by processing method (dry, wet)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Exports by form (e.g. green, roasted, soluble)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Volume of certified coffee exported (e.g. Fair Trade, etc.)					
Quality of exports	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Exports by destination	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

7. How important are the following items relating to coffee imports? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Imports by volume	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Imports by value	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Imports by form (green, roasted, soluble)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Imports by origin	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

8. How important are the following items relating to coffee consumption? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Total consumption of coffee by volume/weight	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Per capita consumption	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Consumption of coffee by type (for example, roasted or soluble)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Single serve consumption (coffee pods, sachets, monodose, etc.)	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Volume of out-of-home consumption	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Coffee consumption by age groups	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Coffee consumption by income levels	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

9. With regard to production, exports, imports and consumption, how frequently would you want data to be made available to you? (Tick all that apply) *

- a. Annually (e.g., calendar year, coffee year, crop year)
- b. Quarterly
- c. Monthly

10. How important are the following items relating to coffee prices? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Coffee prices paid to growers	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
National wholesale coffee prices	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
National retail coffee prices	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
National coffee prices at out-of-home establishments	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

11. How important are the following items relating to coffee stocks and inventories? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Volume of coffee stocks held in exporting countries	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Volume of coffee inventories held in importing countries	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

12. How important are the following items relating to socioeconomics of the coffee sector? *

	Not at all important	Slightly important	Moderately important	Very important	Extremely important
Number of coffee farmers or household	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Employment in the coffee sector	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Coffee sector employment by gender	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Coffee sector employment by age group	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Average size of a coffee farm	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Cost of Production	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5
Coffee farming profitability	<input type="radio"/> 1	<input type="radio"/> 2	<input type="radio"/> 3	<input type="radio"/> 4	<input type="radio"/> 5

13. Please list up to 3 variables, that you believe are important for the ICO to collect, but are not included above:

14. Any other comments?

▲

▼

◀ ▶

RANKING OF VARIABLES FROM SURVEY

Variable	Is this variable currently collected by the ICO	Weighted Score
Cost of production	No	4.52
Coffee prices paid to growers	Yes	4.49
Coffee farming profitability	No	4.45
Exports by volume	Yes	4.44
Exports by type (Arabica, Robusta)	Yes	4.42
Production by volume	Yes	4.40
Production by type (Arabica or Robusta)	Yes	4.39
Total consumption of coffee by volume/weight	Yes	4.39
Imports by origin	Yes	4.34
Exports by destination	Yes	4.30
Quality of coffee produced	No	4.28
Total coffee production area	Yes	4.27
Imports by volume	Yes	4.23
Per capita consumption	Yes	4.20
Volume of coffee stocks held in exporting countries	Yes	4.18
National wholesale coffee prices	Yes	4.15
Volume of coffee inventories held in importing countries	Yes	4.15
Exports by value	Yes	4.14
Number of coffee farmers or households	No	4.10
Quality of exports	Yes, optional	4.09
Exports by form (e.g. green, roasted, soluble)	Yes	4.05
Consumption of coffee by type (for example, roasted or soluble)	No	4.04
Employment in the coffee sector	No	4.04
National retail coffee prices	Yes	4.03
Average size of a coffee farm	No	4.02
Imports by form (green, roasted, soluble)	Yes	3.98
Volume of certified coffee produced (e.g. Fair Trade, etc.)	No	3.96
Imports by value	Yes	3.95
Volume of certified coffee exported (e.g. Fair Trade, etc.)	Yes, optional	3.93
Volume of out-of-home consumption	No	3.88
National coffee prices for out-of-home consumption	No	3.86
Coffee consumption by age groups	No	3.84
Production by processing method (dry or wet)	Yes	3.80
Exports by processing method (dry, wet)	Yes	3.79
Single serve consumption (coffee pods, sachets, etc.)	No	3.79
Age of coffee trees	Yes	3.79
Number of coffee trees in production	Yes	3.75
Total number of coffee trees	Yes	3.74
Coffee consumption by income levels	No	3.71
Coffee sector employment by age group	No	3.68
Coffee sector employment by gender	No	3.63
Varieties/Cultivars produced (e.g. Geisha, Java, Bourbon)	No	3.56