E

ESS/ALSLAC/13/4

# FAO-OEA/CIE-IICA WORKING GROUP ON AGRICULTURAL AND LIVESTOCK STATISTICS FOR LATIN AMERICA AND THE CARIBBEAN

Twenty-sixth Session

**Port of Spain, Trinidad and Tobago, 5-7 June 2013**

## Response rates to FAO questionnaires and collection issues

**FAO**

## Introduction

The agricultural information system is one of the most important building blocks for the formulation of development plans and policies, aimed at improving the efficiency of agriculture production and distribution; as well as the food availability in the world. Article I of the FAO Constitution states clearly that "The Organization shall collect, analyze, interpret and disseminate information relating to nutrition, food and agriculture". In this regard, FAO has constantly given a great importance to the collection, processing and dissemination of food and agriculture statistics. During the years, the FAO statistics have become a global public good, covering the food, agriculture, forests, fisheries and natural resources statistics for more than 200 countries.

In November 2005, FAO endorsed the Principles Governing International Statistical Activities, developed by the Committee for Coordination of Statistical Activities, and thereby expressed its commitment to the principle that “high quality international statistics, accessible for all, are a fundamental element of global information systems” and to the endeavor of continuously introducing “methodological improvements and systems to manage and improve the quality and transparency of statistics”. The Statistics Division of FAO (ESS) is consistently orienting its collaboration with the member countries to improve the availability and quality of, and access to the data for all users.

This paper has two purposes. It aims first at providing the IICA members with an overview of the problems in FAO data collection, of data availability and to some extent of data quality in the region. The second purpose is to illustrate some possible solutions and trigger a discussion that help identify the major causes of the relatively low regional response rate to questionnaires and the possible solutions to improve data availability and quality.

It is important to recall that the quality of the data disseminated by FAO is strongly dependent on the completeness, accuracy and comparability of the collected national data.

The paper has the following structure. After the introduction, section two describes the issues faced in collecting and processing country data. Section three describes the status of data availability and quality in the region based on the questionnaires received. Section four formulates some hypothesis on the possible causes for low response rates. The fifth section finally proposes some possible solutions for discussion with IICA members.

## General issues and FAO data problems in the regions

The twenty-first century started with new challenges for policy makers and development partners with respect to the role of agriculture seen on the one hand as an engine for growth and poverty reduction in addition to its primary importance for food security. On the other hand, the world is confronted for the first time with natural resources constraints where agriculture has become part of the problem with respect to land degradation, water scarcity and pollution, etc.

The first data related issue is that the increased complexity and articulation of policy issues triggered increased data needs and blurred the boundaries between disciplines and dimensions that were neatly separated in the past. The interrelations of the physical, economic, environmental and social dimensions of agriculture have been developed in-depth in the Global Strategy for the Improvement of Agriculture and Rural Statistics.

The relevant point for this paper is the confirmation, based on FAO questionnaire data, of one of the initial remarks of the Strategy that “many countries, especially in the developing world, lack the capacity to produce and report even the minimum set of agricultural data necessary to monitor national trends or inform the international development debate.”

To conclude on this issue: in front of increased and more complex data needs, countries responses are staggering and data completeness is decreasing in the IICA region. The causes for the decrease in data availability and the possible actions need to be discussed. A more detailed analysis will be provided in section 3 and in the annex.

A second major general issue has to do with recurrent country complaints on the data published by international organizations. Inconsistencies are found between national and international data, and sometimes across international organizations, which are due to the processing and validation activities taking place at international level. The United Nations Statistical Commission (UNSC) has promoted two initiatives in since 2012 on the coordination of Statistical Activities in the UN System and on the need to peer-review the data published internationally. While the first initiative is being pursued and a "Friends of the Chair" Group was formed in 2013, the second one was halted because the disadvantages of adding a new international validation layer have been considered to be higher than the benefits. The issue of international data validation from the countries remains nevertheless a very serious one and it will be addressed in sections 3 and 4 of this paper.

On a more tangible level, FAO Statistics Division encounters a number of problems with data collection and processing.

Data collection problems related mainly to data availability and quality. More in particular, each questionnaire has a significant percentage of non-responses. Incoming data are often incomplete and sometimes inconsistent across data sources and over time. In other cases data are not collected according to international standards and need to be harmonized prior to processing.

Data processing causes a different set of data problems and consists of three main activities: harmonization, imputation and validation. The preliminary harmonization or standardization process where data are converted to a common unit of measure and a common classification is the first cause of the discrepancies observed between national and international data. FAO has addressed the root causes at international classification level and has worked for several years to transform its commodity list in an international standard. The shift from FAO commodity list to the *CPC rev 2.1-Extended* international classification will take place starting in 2013[1]. After harmonizing data, gaps in the basic data are filled in with a number of imputation techniques. Imputation of missing data is a necessary step for all international organizations to be able to compile world or regional aggregates, and derived indicators or analytical reports like the Supply Utilization Accounts and Food Balance Sheets. All datasets go through domain-specific and iterative validation procedures that are however internal to FAO and to few selected experts or users.

The will analyze the possible causes of weak data collection in the region and use them **as a guide to identify statistical capacity development needs.** Data imputation techniques are not the focus of this work. Other IICA papers will provide details on selected domains. The paper means to **highlight an essential flaw in the validation process and propose a solution for discussion in section 5.**

[1] FAO's work with the United Nations Expert Group on Classifications on the evolution of the Central Product Classification (CPC), the Harmonised System (HS) and other classifications will be presented in greater detail in separate paper.

## Data availability and quality in the region: a short description of the data received through FAO questionnaires

FAO Statistics Division sends out seven annual questionnaires on production (Prod), producer prices (PP), land use (Land), pesticides (Pest), fertilizers (Fert), machinery (Mach) and government expenditure. Trade data collection follows a different process where countries send directly their trade data files. Furthermore, two ad hoc questionnaires were sent in 2012 to survey country practices on Cost of Production statistics and on the international classifications adopted at country level. Detailed tables on response rates and completeness by questionnaire, country and sub-region are available in the annex.

This section is divided in two parts. The first part aims at identifying the main patterns and data collection issues, so only selected figures and graphs have been incorporated. The second part gives a systematic overview of each questionnaire. North America (Canada and the United States) has been excluded from most graphs because the two countries are not representative of the issues at stake. Government Expenditure has also been excluded because questionnaires were sent for the first time in 2012.

### Overall patterns and data collection problems

Latin American countries taken as a whole show in 2011 lower response rates compared to the world average for almost all questionnaires except those on current agricultural inputs, namely fertilizers and pesticides (Fig. 1). The situation is however more complex, because of the large differences in the sub-regions.

The disaggregated graph to the right shows that South America fares systematically better than the world average, Central America's countries data are relatively weaker on Land and Machinery, and the Caribbean countries have systematic lower response rates.

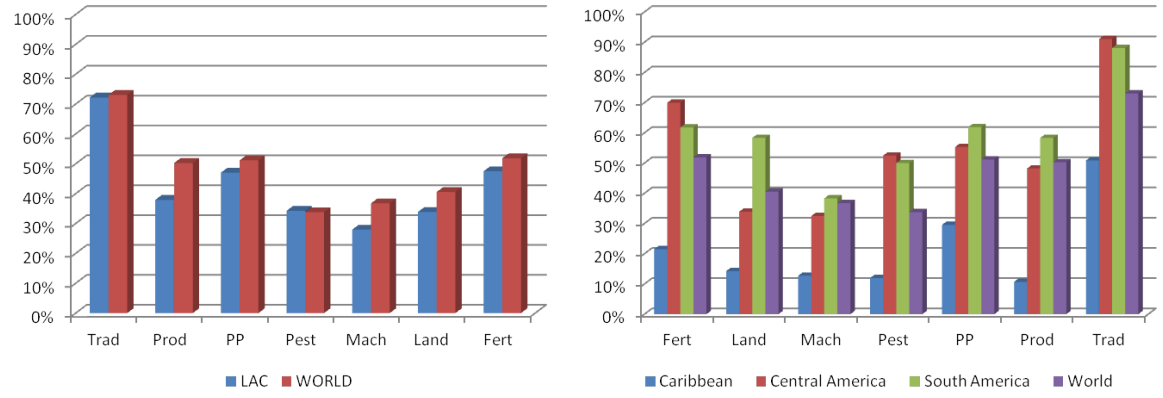


**Fig. 1: Response rates by questionnaire in Latin American and the Caribbean countries and the World, for the whole region and by sub-regions (average response rates 2005-2011)[2]**

More importantly, response rates have declined in recent years for trade, producer prices, land and machinery (Fig. 2-1). The number of responding countries has consistently increased for the production questionnaires and for current inputs. Data collection on fertilizers and pesticides was halted between 2003 and 2007. Fig. 2-2 show that since 2007 the work is consolidating and the information base of the two datasets is increasing.

[2] Machinery (Mach) response rates refer to the period 2005-2010

It would be important to understand the reasons behind these declines, whether they are linked to a decrease in data availability at country level or to other factors. Graphs report the average for the region because similar trends are observed in all the sub-regions.

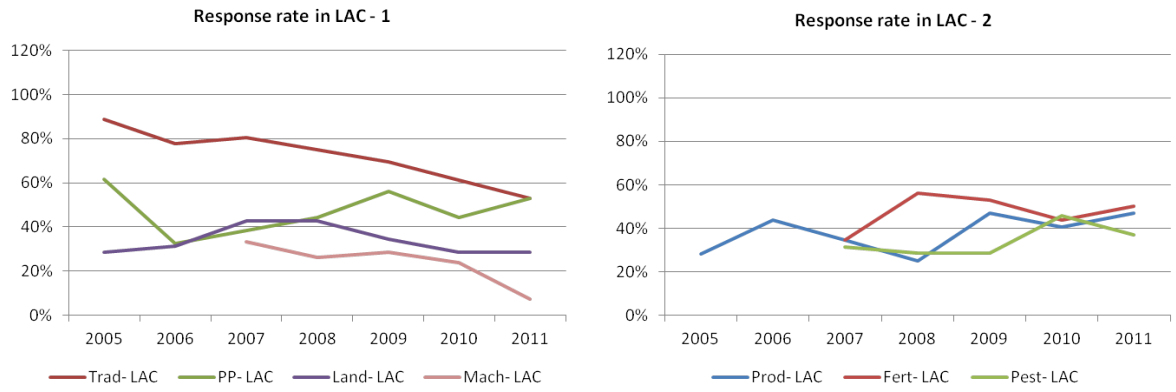


**Fig. 2: Response rates by questionnaire in the LAC region, 2005-2011**

Country response patterns within the sub-regions. In South America even Colombia and Peru, who sent back the largest number of questionnaires could not reply to all seven every year, and a large country like Argentina replies intermittently to the various questionnaires. In Central America, Panama, Guatemala and Belize seem to have many difficulties in providing the data while Mexico, El Salvador, and to some extent Costa Rica respond regularly to all questionnaires. In the Caribbean, only Cuba and St Kitts and Nevis provide up-dates for all dataset with some regularity. Most other countries have difficulties even in reporting on agricultural production.

It would be important to understand the reason for irregular reporting from those countries who have sent more than one questionnaire in the reference period and to identify capacity development needs of those countries who do not have the data at all.

**Data overview for each questionnaire.**

Response rates depend on the kind of data requested. Trade data are the most commonly available with half of the countries having reported every year and average number of responses per country over 5.2, while machinery data are the least available. The questionnaire on producer prices is the second one in terms of average number of responses per country (3.3 in the reference period), followed by fertilizers (3.0), production (2.7), land use (2.3) and pesticides (1.7).

### a) Data collection on Agricultural Production

Agriculture production is a major dataset in FAOSTAT and a key input for the compilation of Supply Utilization Accounts/Food Balance Sheets. The annual questionnaires request production, area harvested, yields, livestock data and selected information for some key processed commodities. Response rates are very high in South America, but do not exceed 50% in Central America and are even lower in the Caribbean. More importantly the questionnaires completeness has weakened in the observed period and number of records has steadily decreased, in all countries but for a few exceptions (notably Brazil, Mexico, Peru). Only 43% of the records are based on official data in 2011 in Central and South America, compared with 60% in 2005. Caribbean countries report a few data in each questionnaire.

The regional workshops of 2010-2011 proved effective in helping countries to provide the data. Should it become a recurrent event or do IICA members have a different suggestion on how to improve production data availability and quality?

### b) Data collection on Agricultural Trade Flows

Trade data collection does not take place through questionnaires. Country send trade data files to FAO and data are processed in collaboration with the UNSD and other organizations. Response rates are falling even on trade. The negative trend affects only two countries in South America (Uruguay and Paraguay) and two countries in Central America (El Salvador and Honduras) but is widespread in the Caribbean's where the number of responses has fallen from 13 in 2005 to 3 in 2011 (Aruba, Barbados and Antigua). This dataset also has the highest percentage of official data, but here again the rate is slightly decreasing in Central and South America and sharply decreasing in the Caribbean when countries report.

It is urgent to identify a strategy in the Caribbean countries to counteract on this trend.

### c) Data collection on Agricultural Producer Prices

The questionnaire on producer prices is filled in by approximately 100 countries in the world every year. The wide availability of price data may be misleading: many countries report wholesale or market prices instead of prices received by the farmers. Response rate is low in the Caribbean along with a weak measure of completeness; while most countries in Central America (except Nicaragua, Guatemala and Belize) and in South America (except Guyana, Venezuela, Paraguay and Uruguay) replies at least every second year. The share of official data keeps decreasing, which deserves attention. Another problem with the producer prices is data inconsistency across time. Series are broken and the new term is greater than the previous one by a factor of 10. In these cases, it is difficult to understand if there was a mistake while converting data, or it is a change in the price concept monitored, or a change in the variety or a combination of all these factors.

IICA members are urged to help understand the causes of the decreasing data coverage and to inform FAO of the actual price concept that they monitor and non-standard units that may be used.

### d) Data collection on Agricultural Inputs

Of the four different questionnaires on agricultural inputs, Pesticides and Fertilizers were halted for a few years and data collection restarted in 2007. The regular information flow is slowly building up. Fertilizers have a high response rate, however their rate of completeness is slowly decreasing and some countries are confronted with data confidentiality constraints when providing data. On the opposite side, few countries respond to the questionnaire on pesticide use. Even Brazil, Uruguay and Argentina could not provide the data more than once in seven years. It must be noted that the share of official records in the dataset is slowly increasing. It would be interesting to understand the reasons for the low response rates and the actual availability of pesticide data at country level.

The land use and machinery questionnaires have in common the lowest response rates and the strongest decrease between 2010 and 2011. Noteworthy, both questionnaires were revised in 2011 to introduce an annex on land use change and new machinery items respectively. Despite a peer-review and the pilot test prior to data collection, results tell that FAO has a lesson to learn from last year. Countries are encouraged to express their views on the matter and formulate suggestions on a better way to manage the introduction of new questionnaires.

It would be important to understand the reasons for irregular reporting: whether data are not collected annually (for example on land use or machinery in use), whether there is capacity issue, or whether the questionnaires need to be improved.

### e) Other data collections

FAO Statistics Division is collecting more datasets one of which, Government Expenditure in Agriculture (GEA) , through a new annual questionnaire since 2012. The Investment Statistics area of work that includes GEA is presented in detail in a separate document.

Other datasets related to the investment in agriculture (Credit to Agriculture, Foreign Direct Investment, External Assistance to Agriculture) are collected from secondary sources in order to avoid over-burdening countries and duplication of efforts. FAO is developing several partnerships with other international organizations with the mandate in the field of interest.

Another new area of work is the development of a Satellite Environmental Economic Accounts for Agriculture (SEEA-Agriculture) that will be presented in detail in a separate paper. The SEEA work includes the development of Economic Accounts of Agriculture. This area of work has just started and FAO Statistics Division is exploring available data sources on-line. The plan however is to initiate data collection on this topic in order to develop a global database on Economic Accounts of Agriculture in the near future.

**Understanding the possible causes underneath incomplete data provision**

The previous paragraphs described the status of data availability and quality to help understand the causes behind it and identify possible ways to improve data collection.

Below is a list of the possible factors affecting responses, on which we would appreciate a feed-back.

a) Questionnaire revisions seem difficult to manage by the countries. Are the changes too disruptive? Do questionnaires come with insufficient explanations?
b) When a country replies every second or third year is it because of a channelling issue but the data exist or are the data not available and the country undertakes ad hoc additional analysis and estimations which are not sustainable on a regular basis?
    - Is data simply not collected on an annual basis?
    - Is it a matter of finding the focal institutions or contact person?
c) Response rates are systematically low in some cases (e.g. machinery and pesticides). Are there different and better ways to collect those data than the annual questionnaires?
d) Where data completeness decreases:
    - Are questionnaires too heavy and long to complete? Is it a matter of resources?
    - Are questionnaires to complex? Is it a matter of capacity?
e) The Caribbean countries have greater difficulties in all respects. Small islands have special structural characteristics (little land, small crops sector) so some data collections are irrelevant for them. Institutionally they are often characterized by small offices with limited number of staff hence low capacity. A sub-regional approach to capacity development should be considered.
f) Data are sometimes conflicts with other national sources or is inconsistent over time.
    - Was there a change in the metadata underneath (different unit, different concept and/or different methodology) that needs to be documented?

**Measures currently undertaken and possible strategies for the future.**

**Current measures**

The insufficient data availability and completeness has an important consequence that deserves attention. FAO Statistics Division is forced to impute missing data to be able compile, derived indicators, regional and world aggregates, or analytical reports such as the

Supply Utilization Accounts / Food Balance Sheets. This is the measure currently undertaken by most international organizations, which creates a distance between the organizations and the countries who disown their own data. Moreover the imputation work is massive and is based on a low information base. FAO has developed a number of imputation techniques for each dataset that are not in the scope of this paper.

The crucial point is that the both aspects of data validation: the peer review of both the methodologies applied and the resulting data do not actively involve the countries. In other words, the current situation is that a) imputation methods are not concurred with the countries and b) imputations are not computed by the countries who could rely on a wider information basis for example on the proxies to use.

- ➔ Should part of capacity development focus on the imputation methods to be used instead of sheer data collection?
- ➔ Can countries be more actively involved in the validation process?

**Short-term solution: enhancing data transmission through greater communication between FAO and countries and strengthening statistical capacities**

FAO Statistics Division devotes a fair share of time in providing feed-back to the countries, up-dating the contact lists and searching additional information. Keeping the dialogue open is demanding on both sides because it requires much staff time.

More concretely FAO would like to make the following request to IICA members:

- ➔ To check and up-date the list of focal points for each country reported in the annex, to help FAO request data to the actual data owners
- ➔ To share comments with FAO on the questionnaires and related instructions, in particular the revised machinery one
- ➔ To inform FAO of the actual data availability and frequency
- ➔ To share the metadata on the actual concepts, definitions and methods used

FAO encourages IICA members to approve the following proposals or formulate different ones with respect to the identification of capacity development needs:

- ➔ Should a second round of regional workshops on FAO datasets be organised?
- ➔ Should new workshops be organized with a focus on imputation techniques?
- ➔ What priority action should be undertaken in the Caribbean sub-region?

**Long-term solution: to establish a platform for peer-reviewing country data published by FAO**

An official instance is needed there countries discuss and validate the methods implemented by the FAO imputed by FAO in the absence of actual information provided directly by the countries. A Committee on Statistics serviced FAO where member countries from all the regions meet could be the appropriate platform to discuss, agree and officially endorse agricultural statistics standards, validate data and decide on priorities.

- ➔ Would IICA members support the creation of a Committee on Statistics?

**Tables**