

Distr.
GENERAL

CES/AC.61/1999/19
23 March 1999

Original: ENGLISH

STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

COMMISSION OF THE EUROPEAN
COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT
(OECD)

FOOD AND AGRICULTURAL
ORGANISATION (FAO)

Meeting on Food and Agricultural
Statistics in Europe

(Geneva, 23-25 June 1999)

JOINT DATA COLLECTION/DATA SHARING BETWEEN
INTERNATIONAL ORGANISATIONS (IO's)

A PROPOSAL TO TEST A NEW APPROACH

Invited paper submitted by the Intersecretariat Working Group
on Agricultural Statistics (IWG.AGRI)*

Summary

Increasing data needs and stagnant or decreasing resources require a new approach with regard to data collection and data dissemination mechanisms. This paper attempts to outline a possible solution, based upon Internet technology. While not universally applicable, such an approach could be tested for specific, well identified fields for agricultural statistics.

* Prepared by Mr. Andreas Lindner, Organisation for Economic Co-operation and Development (OECD).

Introduction

1. Scarcity of resources and budgetary constraints have given new impetus to exploring feasible and satisfactory ways of lowering the statistical response burden for NSOs (**N**ational **S**tatistical **O**ffices) vis-à-vis IO's. In addition, there have been growing pressures to identify and clarify common areas of regular statistical needs across IO's not only to avoid unnecessary duplication in requests to NSOs, but also to reduce discrepancies in published data and to explore scope of joint activities.

2. A number of joint data collections already exist amongst IO's in various fields, such as national accounts, environment, and agriculture. The "Protocol on the Exchange of Statistical Data between Eurostat and OECD", which is under discussion, provides an operational framework with general guiding principles and rules. In this context, great care has to be taken to distinguish between regular and new and/or ad-hoc needs (which may become regular needs later on), and between IO requirements for Member countries and for Non-Member countries (in the field of agriculture they are not necessarily the same), as well as the different time-scales of needs across IO's. Therefore, any sharing agreement has to be looked upon on a **case by case** basis.

3. Concerning Agricultural Statistics, OECD has been asked at the IWG.AGRI meeting in Rome (22 December 1997) to propose a feasible mechanism for joint data collection/data sharing, notably with respect to transition economies. It was understood that the same or a similar mechanism could also apply to the Economic Accounts for Agriculture after OECD's methodological review of this regular core activity (OECD and Eurostat closely co-operate on the EAA without duplication).

4. In early 1998, OECD has circulated the requested proposal amongst IWG.AGRI members and three countries were given the opportunity to provide comments. This note is the slightly re-drafted version of the original proposal.

5. The ideas outlined below go beyond co-ordination practices already in place. Depending on the comments/reaction from countries, it may be envisaged to test it for a particular field in agricultural statistics.

Proposal

6. The following ideas only concern regular statistical activities with a reasonable life expectancy; there is no need to establish co-ordination mechanisms built upon moving sand. The objectives stated are perceived as clear advantages for both the producer and user communities of statistics. The process described attempts to clarify how and under which conditions and mechanisms such an undertaking could be successful. The conclusions derived

point to the feasibility to remove the four main impediments, which have made it difficult so far for IO s to engage in such joint ventures.

Note: Three of the four IWG.AGRI IO's collect agricultural statistics from transition countries: the FAO, Eurostat, and the OECD. The UN/ECE is member of the IWG.AGRI, but not engaged in significant agricultural data collection. However, its co-ordination role in other statistical areas and active involvement in IWG.AGRI work and events would make it desirable that the UN/ECE can access the data of the three other IO's as described below.

I. Objectives

7. Concerning identified **regular** data needs from IO's:

NSOs: provide an identified and agreed range of agricultural data **simultaneously** to all three IO's. This means that no IO receives the earliest available information later from countries than another IO or IOs.

IO's: can simultaneously access their specific part of the above data set at the earliest possible moment in time.

IOs and NSOs can track whether or not changes to data have been made and, if so, which variables have been changed (transparency). To do so, it is imperative to be able to distinguish between INPUT and OUTPUT files.

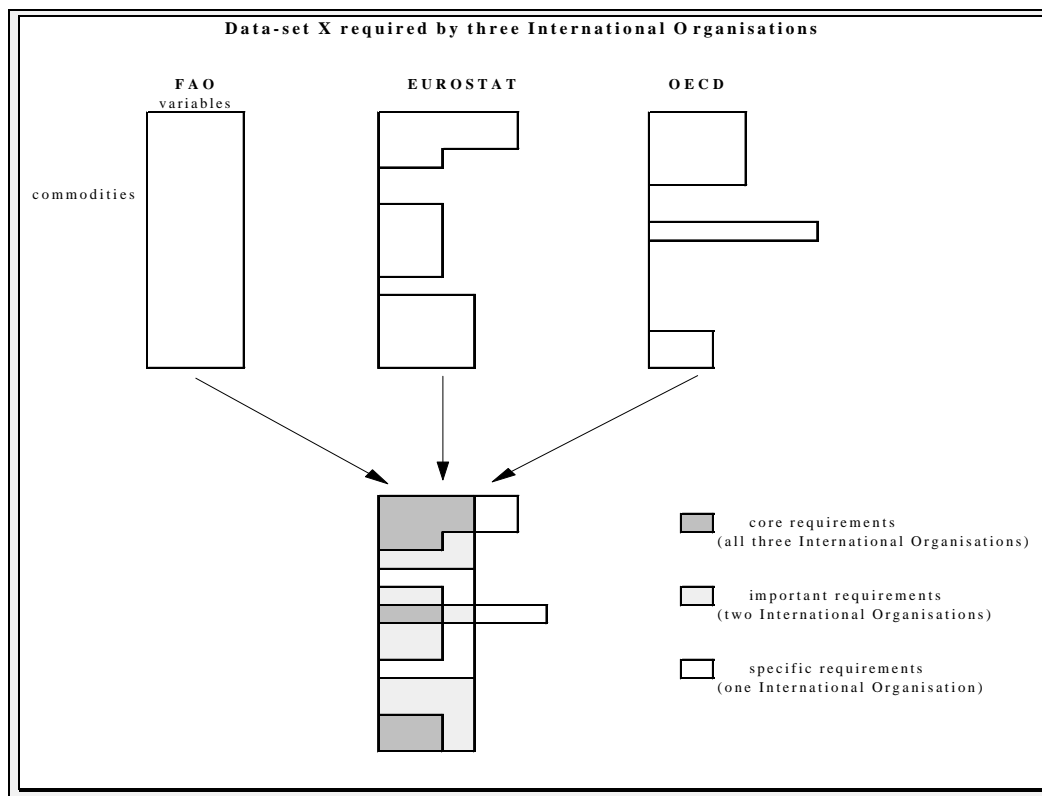
8. The benefits of such a modus operandi are obvious:

- No need to send out separate (or joint) questionnaires.
- No delays in obtaining basic data.
- Transparency amongst all involved as to the basic (=input) data and the IO-specific output data.
- **Internet** will be the medium used (speed)

II. Process

9. IO's identify areas of common interest/data collection activity and merge these into one common framework as illustrated below:

Diagram 1: Identification of needs



The above illustration is hypothetical, but nevertheless shows a **fundamental reality**: OECD's needs are certainly more limited in terms of detail than either FAO's or Eurostat's; but in some instances OECD needs data not covered by the two other IO's (e.g. policy indicators and parameters, subsidy calculations etc.). On the other hand, Eurostat may need more details for specific commodities than the two other IO's in order to respond to specific CAP requirements. The "superposition" of the three IOs needs profiles permits to identify where to start with data coordination: the core requirements of all three IO's. It would be desirable to add to the core requirements those which would meet the needs of two IO's ("important requirements") because it would contribute to reducing the response burden for NSO's. The third category, "specific requirements", would not lend itself for any co-ordination since only one IO requires this data. If desired, it could, however, be useful to include it for information for others.

10. In order to achieve such a data sharing, two considerations are of particular importance. First, the **basic definitions and concepts** used have to be identical or at least compatible and the common framework has to be consistent in terms of **classification list(s)** used. Second, the basic variables have to correspond to a statistical subject matter, for which results are obtained from NSO's at a typical moment during the year. This

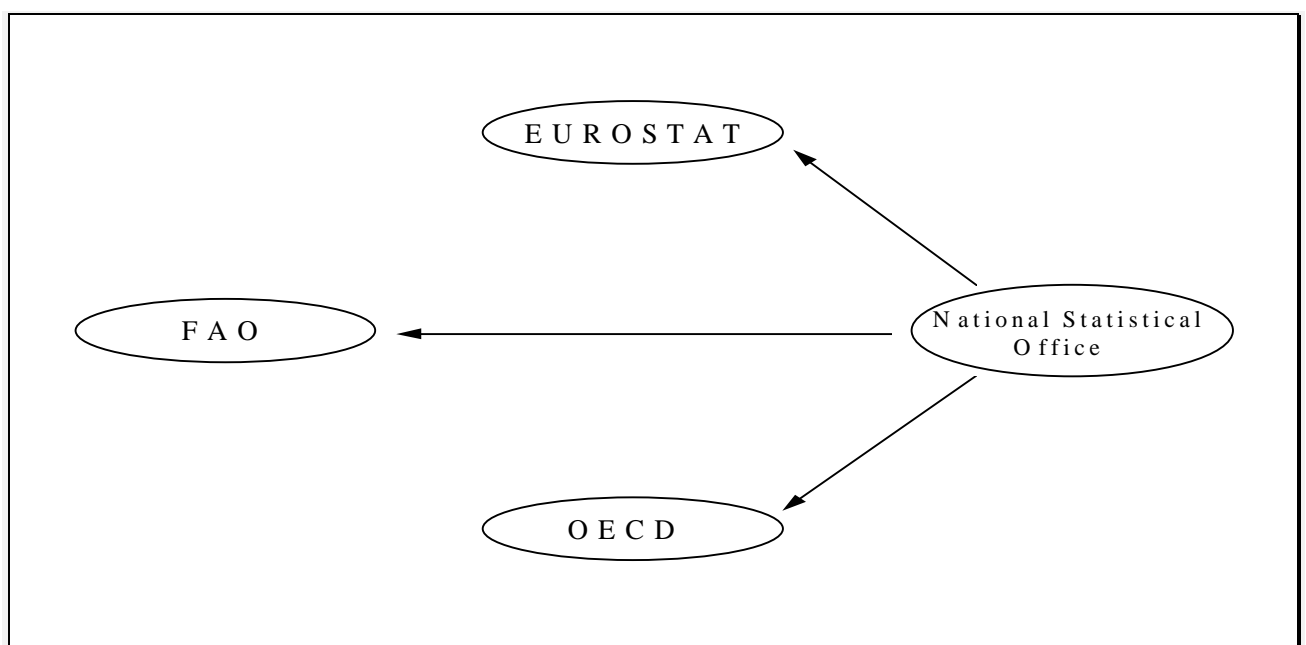
requires to analyze the functioning of national statistical systems, a task which has been carried out by OECD for 11 transition economies and resulted in the optimization of timing of 2 CEECs/NIS agricultural indicators questionnaires per year. The reason for this is not only to avoid asking for information before it can be made available, but also to identify

- which other subject areas could be usefully asked for at the same time and
- to partition the framework into components classified according to availability throughout the year.

11. As in other statistical areas, there exists in agriculture a fairly consistent pattern across countries **when which type of data is collected, compiled and published**. For the sake of timeliness, it may be advantageous to partition and spread over the year the range and degree of detail of statistical subjects collected according to availability. One might sequence data availability by key data available at the earliest possible moment in a given year, followed a couple of months later by the complete data set. Similarly, one might make arrangements that a fairly complete set of "soft" data is available at the earliest possible moment while ensuring that it will be replaced by "hard" data as early as possible.

12. An agreement would have to be reached with NSO's to provide (available) data of common interest to three or two IO's using a standard file format (the INPUT - file). The way this could work is illustrated below:

Diagram 2: The INPUT data flow



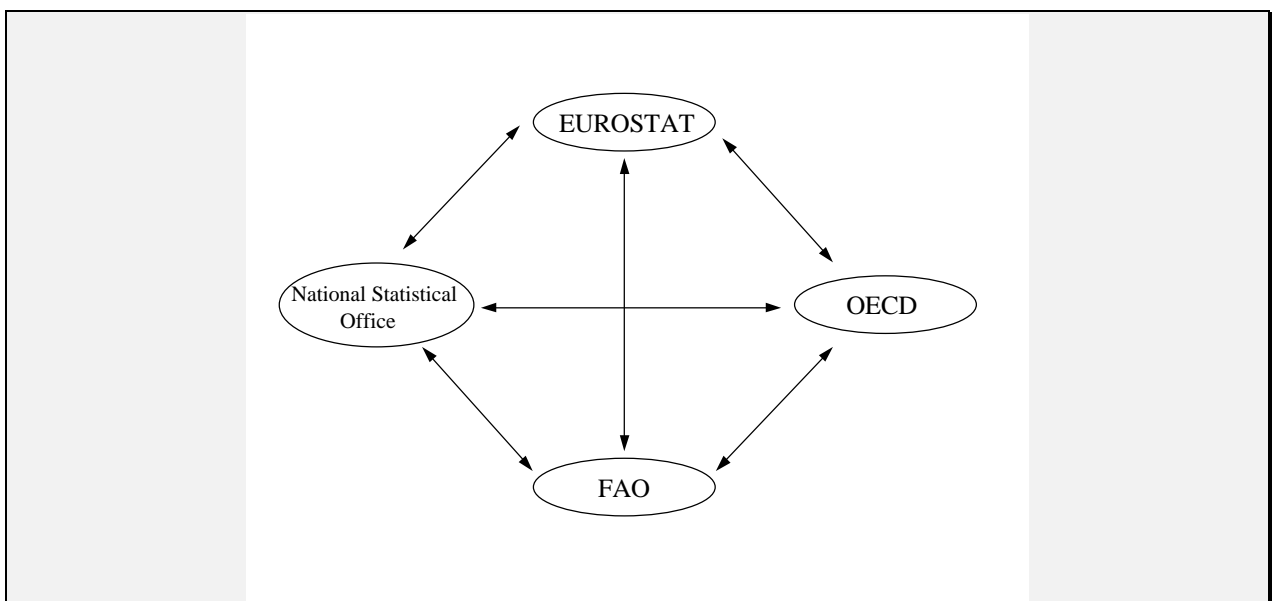
The diagram above illustrates that a NSO **sends, as soon as the data can be officially released, simultaneously and** via **Internet** to all three IO's an **Input file** which is of common interest to at least two IO's.

13. The format (e.g. Excel spreadsheets) and layout should be standardized. Also **metadata** should be included or attached. It seems recommendable that the NSO sends the Input file via Internet to the appropriate Internet addresses and at the same time informs through e-mail the interested service in each IO that this has been done. The reason for this is that the Internet address and the interested service are most often not identical.

14. The next step consists of the extraction, checking and analysis by each IO of the part of the Input file they are interested in. This process also includes possible corrections/adjustments of data (to be included in a metadata file and made known to the NSO concerned and the other IOs) and the use of this modified data for IO - specific statistical products that may differ from those of other IO's. Examples of differences include IO - specific calculation routines, conversion factors used for transforming raw data into other entities, the addition of specific variables and also re-formatting of data files using IO - specific classifications. An important element in this step is the necessary **bilateral follow -up** between each IO and the NSO concerned on specific methodological (or other) questions. This direct, interactive relationship between user and provider is an aspect that can not be left to another IO.

15. For the sake of full transparency and to facilitate methodological comparisons, it would be useful if the IO's would use the same procedure as described above to make their respective OUTPUT- files available via **Internet** (see below).

Diagram 3: The OUTPUT data flow



16. In case that the IO OUTPUT file is different from the INPUT file, there would be no ambiguity to which factor this discrepancy is attributable; the other agencies can easily track this. This case would need extensive annotations and methodological notes to explain in detail why original data had to be modified.

17. In case that an IO - OUTPUT file corresponds exactly to the IO's - INPUT file **but not to the IO's own statistical output** (= different structure, additional variables, coefficients, etc.), the sharing of the respective OUTPUT - files has a number of additional advantages:

First, it would show to other IO's and the NSO concerned why and how some national data had to be changed through adjustments **not attributable** to raw data but rather to the IO's specific way of calculating and publishing figures.

Second, it would help IO's and the NSO to understand why a finalized statistical output of one IO may have different figures: either the cause is a modified input data value or the cause lies in the subsequent, IO - specific treatment which is unrelated to the INPUT file. The possibility of checking on both INPUT and OUTPUT files allows an observer to identify the origin of apparent discrepancies.

Third, it is an advantage to compare the own OUTPUT file with one or two other IO's concerning common elements. This might lead to further harmonization of data disseminated by IO's.

III. Conclusion

18. Provided that the conditions stipulated are fulfilled, the process outlined above should have a good chance to meet the stated objectives because it could remedy to the **four main causes** for continuing individual data collection by IO's for essentially the same data:

Different needs: IO's have an incentive to standardize and/or unify classification lists. The argument of different data structures would not be applicable within the scope of the proposed process, since excluded from the scope of the proposal.

Factor time (input): Since the NSO would share survey results simultaneously with all interested IO's at the earliest possible moment in a given year, there would be no advantage for any IO to launch an individual questionnaire before that release date.

Factor time (output): The IO's processing, checking and validating of national data takes time which may be considered too long for the needs of another IO which depends on another IO for a particular set of data. The

modus operandi described above would allow each IO to immediately work on the input data and to share with other IO's and the NSO the results once they become available.

Data source: No IO can afford in the long run to be excluded from direct contacts with the original data providers. Both data quality and expertise would vanish over time. Since the above proposal would not present an additional burden to NSO's, but rather lead to an increased awareness of different needs and concerns expressed by IO's, direct bilateral contacts can be considered as a natural ingredient of international statistical activities with obvious advantages. In addition, the suggested transparent and open approach puts each IO at the same level of contact with the NSO, while respecting (through decoupling) the IO - specific needs.
