FAOSTAT2 Project: Revising the Conceptual Statistical Frameworks, Underlying Statistical Methodologies, Database System and Processes

Robert Mayo, Statistics Division
Food and Agriculture Organization of the United Nations

Abstract

The modernization of FAOSTAT (the major FAO statistical database) will provide improved methodologies and data quality, streamlined system processes, improved user access to statistical data and a stable and reliable technical environment. This paper outlines the aims and structure of the FAOSTAT Project and introduces the sub-project, CountrySTAT, which will be developed in parallel to FAOSTAT.

The FAOSTAT Project will provide internal statistical frameworks and coordinating mechanisms for ensuring coordination and consistency of statistical methodologies and policies in FAO throughout the statistical process. This will include the development of integrated conceptual statistical frameworks for the major substantive domains such as: Food (Production, Trade, Population, Food Security); Resource (Inputs and Production); and Economic (Inputs, Production, Prices). The structure of the system will also allow for satellite statistical modules to be included in the overall framework. Vertical (country) and horizontal (thematic) integration of statistical data in the FAOSTAT Family will provide increased value to the FAOSTAT statistical system.

CountrySTAT, a scaled-down version of FAOSTAT will provide satellite statistical modules for countries to implement as required. The major objectives of the CountrySTAT project are statistical capacity building; improved statistical data, and the facilitation of data use by national policy-makers. CountrySTAT will assist countries to develop a statistical information system containing available data and metadata relevant to agricultural policy and allow it to be linked (vertically) to the FAOSTAT Family.

Background to the Modernization of FAOSTAT

The FAOSTAT statistical system is one of FAO's most important corporate systems. It is a major component of FAO's information system, contributing to the Organization's strategic objective of collecting, analyzing, interpreting and disseminating information relating to food, agriculture and nutrition. FAOSTAT is a well known product throughout the United Nations, statistical and academic worlds. Policy formulators, decision-makers and other stakeholders, both at the national and international levels are the primary users of the system.

The FAOSTAT working system (i.e. the underlying system used to compile, validate, transform and analyze statistical data) has been operational for over a decade. In recent years, the technical and functional limitations of the working system and dissemination have become more apparent, especially given growing user expectations. Work is now proceeding on the modernization of the FAOSTAT system. In particular, it has been recognized that data quality in the current FAOSTAT system suffers from the lack of a

structured coherent and consolidated data or outdated statistical frameworks in some of the key statistical domains. The modernization of FAOSTAT has provided the opportunity to revisit the underlying data, database and dissemination structures and developing new structures that will provide enhanced data quality and sustainability of the FAOSTAT system.

Structured frameworks for international statistical databases

The modernization of the FAOSTAT statistical system provides a unique opportunity to review and enhance the whole statistical process (see Figure 1) from the collection of data at national level until the dissemination of data and metadata to the users.

Figure 1. FAOSTAT Data Flow
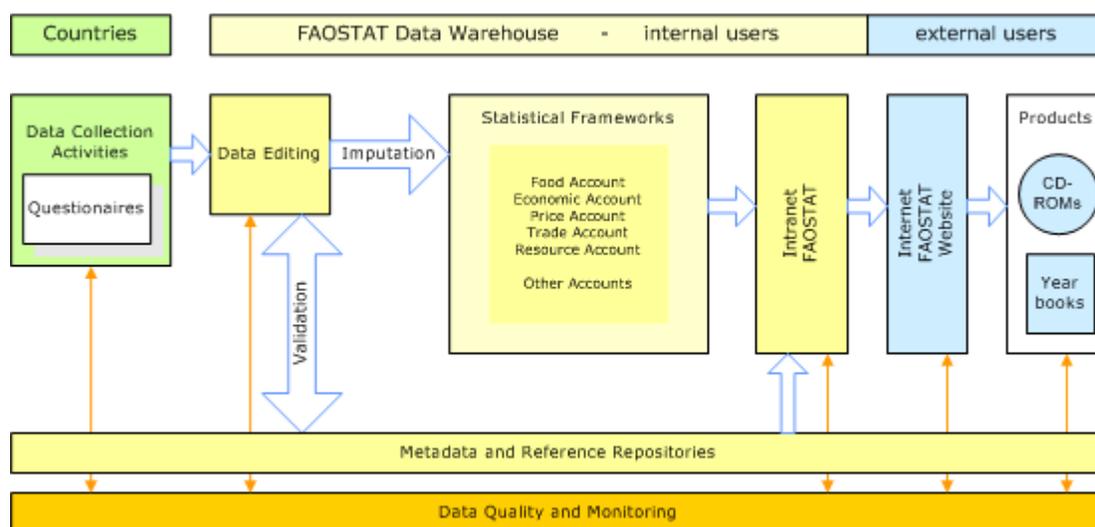
The new FAOSTAT system revolves around a core FAOSTAT module (see Figure 2) with distributed database modules around the core module. This model provides a flexible approach as the satellite databases need only to have linkages to the core and other modules to enable data interchange. The core module will have standard statistical metadata elements to facilitate data interchange with the other database modules. Only selected statistical data will be included in the core module (see Figures 3 and 4). The structure of the system will also allow for satellite statistical modules to be included in the overall framework and various indicators to be developed or included as required.

The new system (via CountrySTAT – a scaled down country-based version of FAOSTAT) will provide the capacity to store and report on country data that is captured at the sub-national or administrative unit level.

Figure 2.   FAOSTAT Family - Core and Related Data Modules
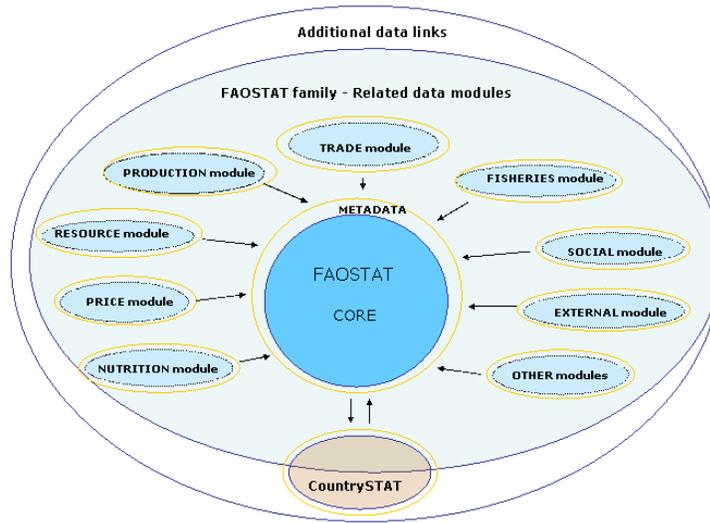
The new FAOSTAT system will provide internal statistical frameworks and coordinating mechanisms for ensuring that statistical methodologies and policies are applied consistently throughout the statistical process. This will include the development of integrated core conceptual statistical frameworks for the major substantive domains such as: Food (Production, Trade, Population, Food Security); Resources (Inputs and Trade); and Economic (Inputs, Production, Prices). See Figure 3 for an overview of the initial core FAOSTAT Framework.

Figure 3.   The new FAOSTAT Core Statistical Framework
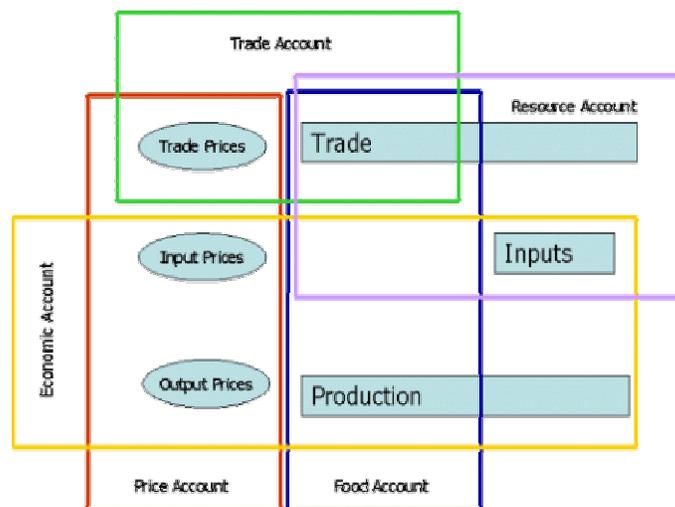
Each of the core statistical frameworks will provide a structured format for the statistical data. From these structured formats, key policy indicators will be produced. The Core Food Module (see Figure 4) provides an example of how this approach will be implemented, taking initial agricultural production and trade data and structuring it and producing estimates on undernourishment. The FAOSTAT core frameworks will provide a consistent set of data for further analysis.

Figure 4.   FAOSTAT Core Food Module
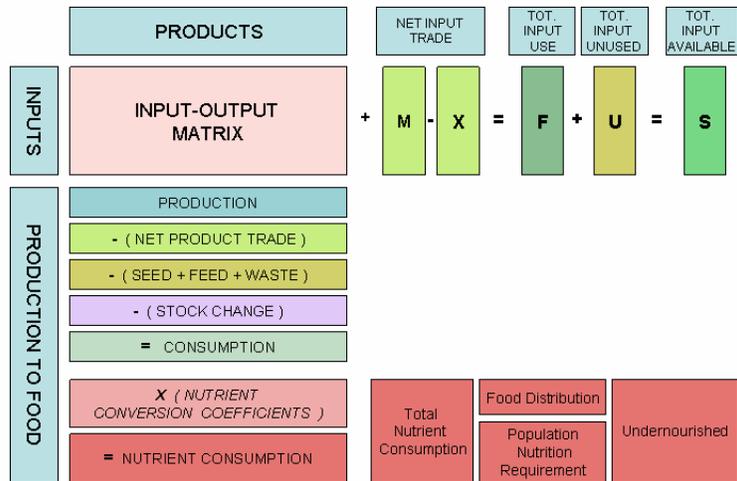
The FAOSTAT Core:

The FAOSTAT Core is a collection of consolidated, coherent and complete datasets based primarily on official data collections providing the essential datasets required for agricultural policy formation and analysis. The data to be include in the Core will need to meet criteria such as: time-series format (annual calendar year); global coverage using statistical recognized boundaries; coherent (meaning that there are no missing values - when necessary values will be estimated to fill the gaps). The Core will include approximately 150 commodities (the Supply Utilization Account (SUA) commodities) from approximately 200 countries and 100 dimensions for all years. The FAOSTAT core will provide a complete dataset in which users will benefit from a balanced and complete data set for analysis.   The data which is not in the Core will be kept in Related Data Modules and Additional Data Links

FAOSTAT - Related data Modules:

Datasets directly related to the FAOSTAT- Core datasets but not part of the essential set necessary for agriculture policy formulation and analysis at macroeconomic level are included in the FAOSTAT- Related Modules. These Related Modules (production, trade, prices, resources, etc) would house the detailed FAOSTAT data and would not include estimated or calculated data.

FAOSTAT- Additional Data Links:

This group is composed of all the datasets and databases that do not meet the criteria to be included in the FAOSTAT family (Core and Related Data Warehouse). Thus, data sets that are of value (thematically or detail wise) to the data in FAOSTAT but may have different time dimensions, statistical boundaries or structures will be linked to the FAOSTAT family.

Horizontal (thematic) integration of statistical data at international level - the FAOSTAT family

The integration of data in FAOSTAT will provide considerably greater value to users. One of the main activities to be carried out to improve the utility of the data component of FAOSTAT is to review the structure and quality of current datasets and analyze possible new datasets that may be available for dissemination. This will include both internal datasets and those from external sources. Delivery of these new datasets will make available important information required by FAO users for data analysis.
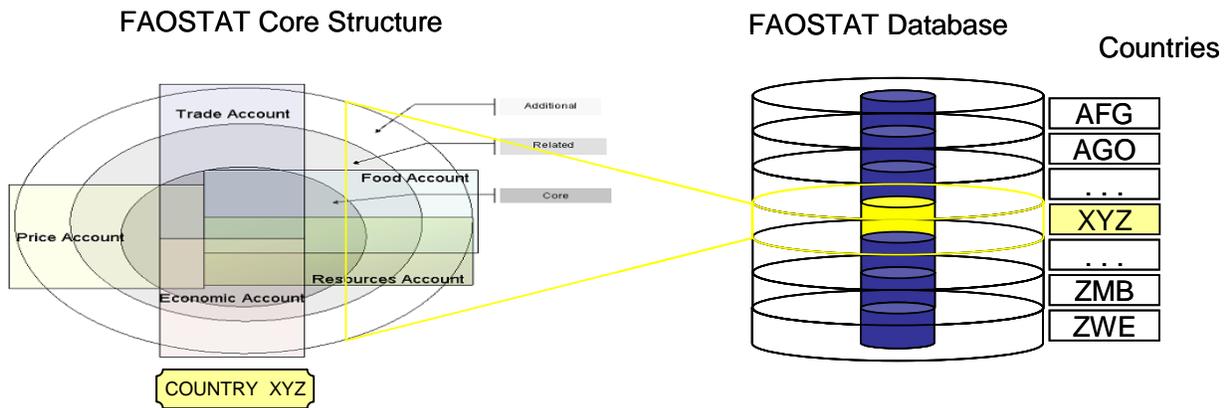
Figure 4.   Horizontal Integration of FAOSTAT Thematic Modules
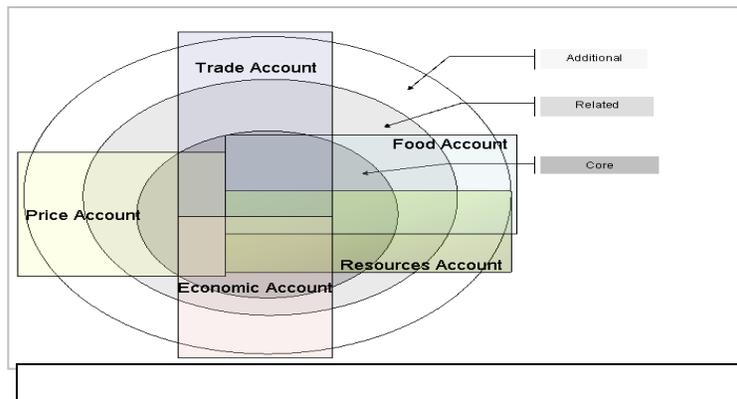
Horizontal and vertical (country) integration of statistical data at international level

As can be seen from Figure 5, the FAOSTAT data warehouse will provide horizontal and vertical integration of data. The Statistical Metadata is the key to linking data at both horizontally and vertically in FAOSTAT.
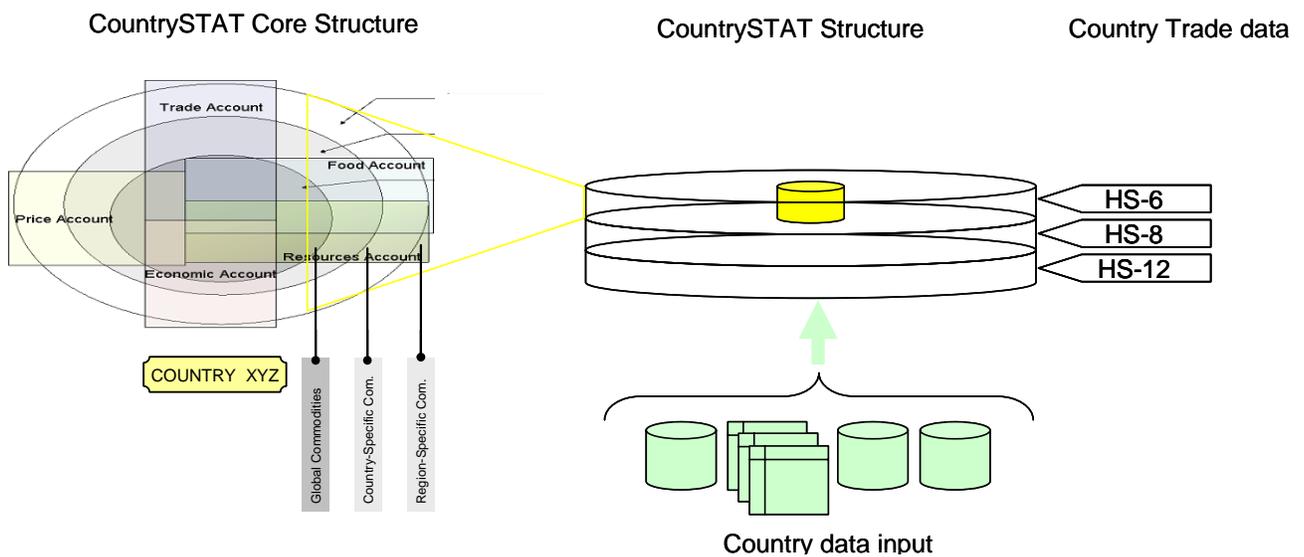
Figure 5.   Horizontal and Vertical Integration of FAOSTAT Thematic Modules

FAOSTAT Core Structure

FAOSTAT Database

Countries

Trade Account

Additional

Related

Food Account

Core

Price Account

Resources Account

Economic Account

COUNTRY XYZ

AFG
AGO
. . .
XYZ
. . .
ZMB
ZWE

Horizontal and vertical integration of statistical data at national level

The country version of FAOSTAT - CountrySTAT will follow the same data structure as FAOSTAT, but will be broader and deeper – i.e. greater horizontal and vertical dimensions/depth will be included, thus adding greater detail to the data. As can be seen from Figure 6, the same Core structure will be implemented but have detail such as country specific detail for international trade statistics. International trade merchandise trade statistics are used in Figure 6 to show the structured frameworks for vertical and horizontal integration using the 6-digit, 8-digit and 12-digit Harmonized System (HS) based coding. This allows for greater detail to be provided (depth and breath) than in FAOSTAT.

Figure 6.   Horizontal and Vertical Integration of CountrySTAT Thematic Modules

CountrySTAT Core Structure

CountrySTAT Structure

Country Trade data

Trade Account

Food Account

Price Account

Resources Account

Economic Account

COUNTRY XYZ

Global Commodities

Country-Specific Com.

Region-Specific Com.

HS-6
HS-8
HS-12

Country data input

The CountrySTAT database will allow for broader datasets than those included in FAOSTAT. This will allow countries to include data specifically relevant to their own needs but consolidated in the CountrySTAT Core and to provide integration with other data from their country that would not normally be in a format for analysis. For example, the use of a consistent CountrySTAT Core will allow Country X to directly compare data with Country Y.

Vertical integration of statistical data between national and International levels

The vertical integration of statistical data between CountrySTAT (countries) and FAOSTAT (FAO) will be possible due to the use of consistent data concepts, classifications and structures and other metadata. As can be seen from Figure 7, data will be able to be exchanged in either direction and many possibilities will available for adding value to data by increasing its accessibility for analysis.

Figure 7.   Vertical integration of statistical data between national and International levels
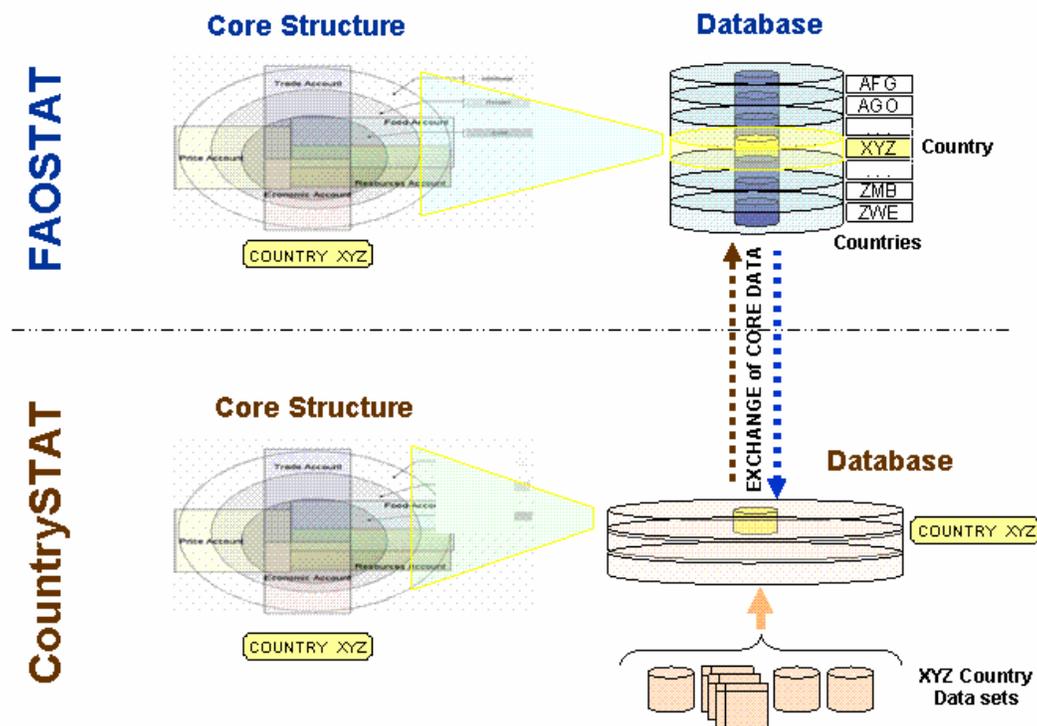
Key ingredients in statistical data integration - methodologies and data standards

The FAOSTAT system relies on the inclusion of data from diverse sources to address the needs of more organizational units within FAO and outside FAO. By conforming to common data concepts, definitions, codes and methodologies. The new FAOSTAT is revisiting existing methodologies, identifying areas for improvement which include standardization of data concepts and definitions. The requirement analysis recognized the need to address data management and statistical methodology issues. This includes identifying the appropriate organizational mechanisms to support ongoing data management and statistical methodology review.

Statistical Metadata

Metadata enhances understanding of any given data item within the system by documenting its definition, history of its values, methodology used in its collection, national contacts, etc. Statistical metadata is one of the keys for the integration of national and international statistical data in FAOSTAT. This information is also useful to statisticians who compile, validate and analyze the data, as well as to the users, both internal and external to FAO, who access the data. The construction of a metadata repository and its integration with FAOSTAT statistical data are essential components of the new FAOSTAT system. An overview of the new FAOSTAT Metadata System is provided in Figure 8.

Figure 5. Overview of the new FAOSTAT Metadata System
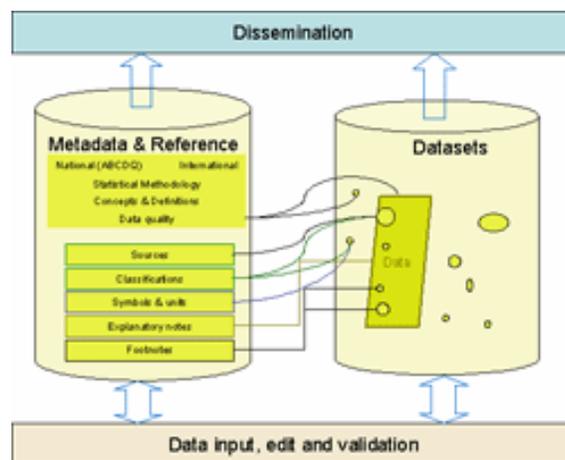
A central repository and standardized methods and modules for documenting metadata will be developed in order to avoid redundant collection and maintenance of metadata, as well as inconsistencies in the statistical process. The statistical metadata system will cover: concepts and definitions; classifications; symbols and units; explanatory notes; statistical methodology; data dissemination; data and metadata quality.

Data Quality Integration

FAOSTAT data quality ultimately depends on the quality of the data reported by member countries and processed by FAO. FAOSTAT is developing standard data quality evaluation and monitoring processes of all statistical datasets to be included in the FAOSTAT Family. Preliminary work has begun on developing a description of the evaluation and monitoring[1]. The standard data quality evaluation and monitoring processes covers both data submitted by countries and data and then processed within FAO. On the basis of the quality descriptions, a list of process-oriented data quality indicators will be identified.

A predefined format (prepared by FAOSTAT) will identify the minimum and standardized set of information items to be included. Data quality - including value added and performance - will be monitored at three different points of the statistical process. Data quality problems identified at any point can be addressed directly, often in real time. Feedback loops will provide a mechanism for improving data quality (see Figure 9).

Figure 9.   FAOSTAT Data Quality monitoring and feedback system
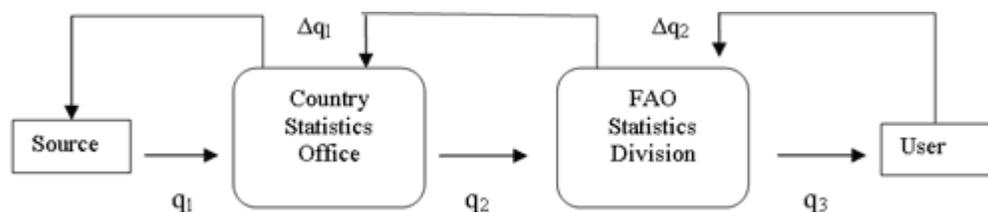
Quality indicators are compatible with those used by other agencies in the International Statistical System, adjusted to the specifics of agricultural data. FAO follows closely the work being carried out by other international statistical offices on data quality monitoring.

Data dissemination

The FAOSTAT data portal is the most visible worldwide product offering users (policy formulators, decision-makers, statisticians, journalists, etc) a fast and reliable way to access FAO statistical information in the areas of food and agriculture. The new FAOSTAT structured statistical frameworks will greatly enhance the functionality of the portal and will strengthen the capacity of users to perform more substantive analytical work such as cross-domain data selection and analysis.

An improved interface, with the capacity to visualize FAOSTAT data and metadata or download them in different formats for further use, will provide users with their preferred tools for analysis. Key country indicators in the form of precompiled reports will provide a useful vehicle to assess a situation at a glance.

---

[1] http://unstats.un.org/unsd/accsub/2004docs-CDQIO/1-FAO.pdf

References

Food and Agriculture Organization of the United Nations, *FAOSTAT2 Project,* 2004. <http://www.fao.org/faostat/>

Food and Agriculture Organization of the United Nations, *CountrySTAT,* 2004. <http://www.fao.org/faostat/countrystat_en.asp>

Food and Agriculture Organization of the United Nations, *Final Report - Informal Expert Consultation on CountrySTAT*, 2003. <http://www.fao.org/es/ess/meetings/fin_rep.asp>

Food and Agriculture Organization of the United Nations, *Final Report - Informal Expert Consultation on CountrySTAT*, 2003. <http://www.fao.org/es/ess/meetings/fin_rep.asp>

Food and Agriculture Organization of the United Nations, H. Kasnakoglu and R. Mayo *FAO Statistical Data Quality Framework: A multi-layered approach to monitoring and assessment,* Conference on Data Quality for International Organizations, Wiesbaden/Germany, 27-28 May 2004, http://unstats.un.org/unsd/accsub/2004docs-CDQIO/1-FAO.pdf