September 2005

Agenda Item 6.1

AFRICAN COMMISSION ON AGRICULTURAL STATISTICS

Nineteenth Session

MAPUTO, MOZAMBIQUE, 24 - 27 OCTOBER 2005

The redevelopment of the FAOSTAT statistical system

Table of Contents

Table of Figures

Background

This paper summarizes the major activities undertaken in the redevelopment of the FAOSTAT[1] statistical system. 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. It lies at the core of FAO's World Agricultural Information Centre (WAICENT) through which access is given to FAO's vast store of information on agricultural and food topics – statistical data, documents, books, images, and maps.

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. They regularly consult FAOSTAT as a data source for analysis and decision-making. Other users of the FAOSTAT system include FAO staff, the international community, researchers, private enterprise and the public at large. It is estimated that approximately one quarter of all visits to the FAO Web pages are made with the purpose of retrieving statistical data. FAOSTAT supports a subscriber base allowing users to perform bulk data downloads for analytical purposes. Data contained in the FAOSTAT system are published regularly, both in hard copy yearbooks as well as on CD-ROMs.

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, its technical and functional limitations have become more apparent, especially given growing user expectations. In late 2001, a proposal to proceed with a requirement analysis for the modernization of the FAOSTAT working system was endorsed and work is now proceeding on the modernization of the FAOSTAT working system. The scope of the FAOSTAT Project has changed from the initial focus on the working and dissemination systems to a full re-development of fundamental statistical methodological frameworks and statistics systems that are the foundations of FAOSTAT.

Objectives of the New FAOSTAT System

The improved system will:

Provide tools for compiling, validating, estimating and analyzing data;

Revise statistical methodologies and ensure data quality standards are applied consistently;

Develop a centralised statistical metadata system;

Enhance and create mechanisms for improved user access to data;

Streamline system processes and provide a stable and reliable technical environment.

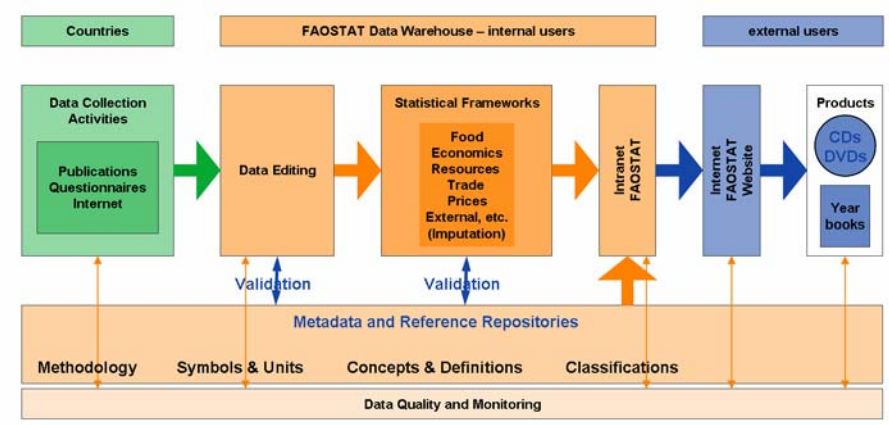
The development of CountrySTAT, a scaled down version of FAOSTAT has provided a mechanism to integrate data, metadata and methodologies from various sources to assist national agricultural statistical systems.

The FAOSTAT Statistical System

An overview of the new FAOSTAT data flow (see Figure 1) from national data to end users shows the scope of the modernization of the FAOSTAT statistical system.

[1] http://www.fao.org/faostat

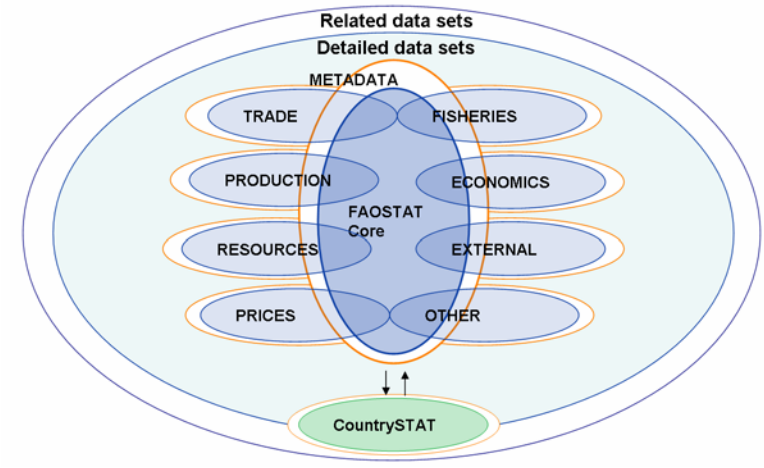
Figure 1. Data Flow in the FAOSTAT Statistical System



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) will provide the capacity to store and report on country data that is captured at the sub-national or administrative unit level. CountrySTAT is a sub-project under FAOSTAT and its development is seen primarily as a by-product of the development of FAOSTAT. It is intended to take a stepwise approach to CountrySTAT with pilot studies of the core CountrySTAT and the various modules being undertaken in selected countries at different stages of development.

Figure 2. FAOSTAT Core and Satellite Database System

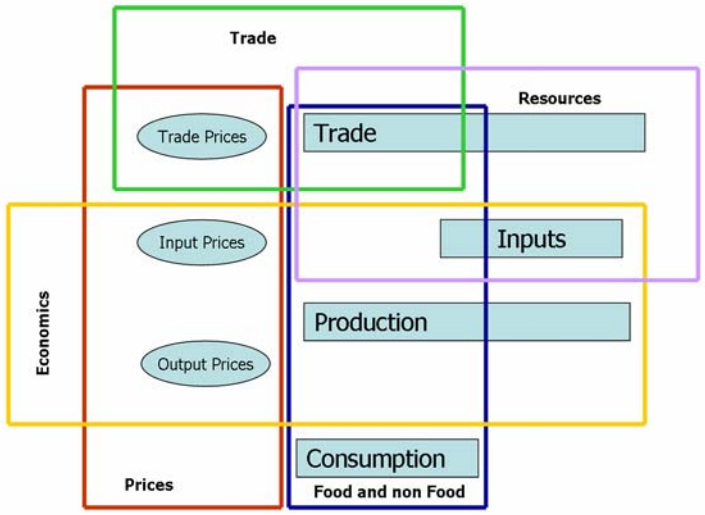


Statistical Domains and Frameworks

Integrated statistical frameworks are being developed for the major substantive domains of **Food** (Production, Trade, Inputs, Consumption); **Resources** (Inputs and Trade); and **Economic** (Inputs, Input Prices, Production, Output Prices) **Prices** (Input Prices, Output Prices, Trade Prices), **Trade** (Imports, Exports, Prices). The integrated statistical frameworks constitute the core consolidated data in FAOSTAT. The statistical

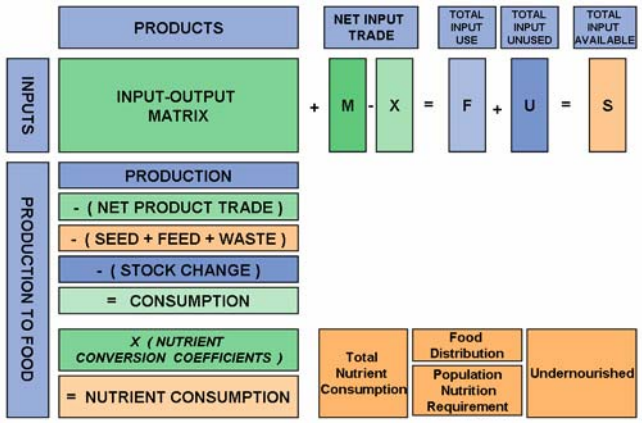
frameworks are made up Core Consolidated detailed datasets. By having the detailed datasets as satellites to the Core Consolidated data, additional datasets can be included in the overall framework as required (see figure 3).

Figure 3. FAOSTAT Core Statistical Frameworks



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.

Figure 4. Core Food Framework



Statistical methodologies

A peer reviewed recognised methodology framework for estimating and balancing and thus producing a consistent coherent dataset within the statistical framework domains (Food, Prices, etc) has been developed and tested. These statistical methodologies will be used to generate datasets for the further generation of key statistical indicators in FAOSTAT such as for under-nourishment indicators. This new methodology improves data quality throughout the FAOSTAT system and improves consistency of data by ensuring a standardised approach to the balancing of data is applied.

Metadata

Structured and comprehensive and centralised metadata is a major new feature of FAOSTAT. The metadata system provides a framework for managing FAOSTAT metadata, thus providing procedures and mechanisms to avoid redundant collection and maintenance of metadata, resolve inconsistencies and allow for easy user access to many layers of metadata for each data item. A central repository and standardised methods and modules for documenting metadata have been developed (see figure 1.). FAOSTAT actively collaborates with statistical colleagues in the Statistical Data and Metadata Exchange (SDMX) initiative[2] and has mapped terminology in the FAOSTAT Metadata system with the Metadata Common Vocabulary (MCV).

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. This information is 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. The statistical metadata system will cover: concepts and definitions; classifications; symbols and units; explanatory notes; statistical methodology; data dissemination; data and metadata quality.

The efforts draw on substantial work dedicated to building the FAO's ABCDQ[3] system and on efforts in the international statistical metadata community. The FAOSTAT project will coordinate the initial collection and organising of the metadata in collaboration with relevant substantive units.

Figure 5. FAOSTAT Metadata Glossary

Classification

The current FAOSTAT list of commodities is a list of agricultural products that serves the specific needs of agricultural production statistics. The list was inspired by SITC and adapted by FAO where needed; in most cases additional detail was added to SITC list. FAO held an Expert Group Meeting on Classifications in Agriculture in August 2004 in which it was recommended that the FAOSTAT commodities be reclassified according to commonly used international classifications. Following the recommendation of this meeting, the project team elaborated a Harmonized System-based approach for the FAOSTAT commodities list. Since the HS, CPC, SITC and the FAOSTAT lists are related, comparing CPC and the FAOSTAT list through HS, which provides the building blocks of CPC, is possible. The current FAOSTAT country codes and groupings are being

[2] http://www.sdmx.org/

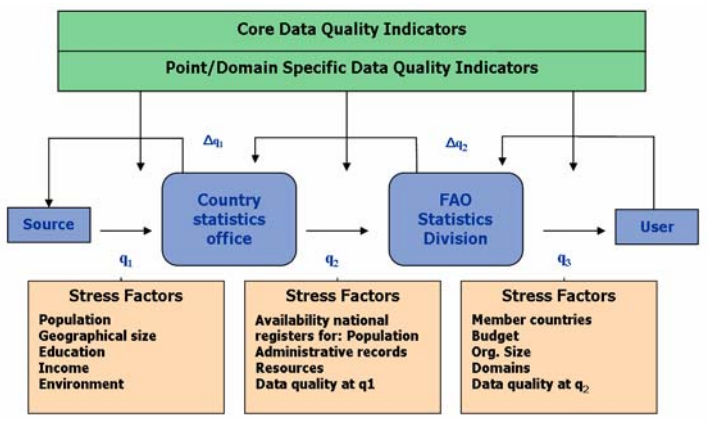
[3] http://faostat.fao.org/abcdq/about.htmx

supplemented by the UNSD M49 country codes and will allow users to download data with their preference of country codes.

Data Quality

FAOSTAT is developing a system which aims to monitor data quality in all aspects of the international data process (from Country to FAO dissemination). The key aspect to this approach to data quality evaluation and monitoring is the focus on three different points of the statistical process (when it enters the national office, when it leaves the national office, and when it is disseminated by FAO). Currently international organisations focus on usually just one point of the statistical process and therefore do not take full advantage of information on data quality available at various points of the statistical process (see figure 6). By monitoring data quality at the three key points it is possible to assess value-added by each stage of the process and assess performance. On the basis of the quality descriptions, a of list data quality indicators are being identified for each stage of the process.

Figure 6. FAOSTAT Data Quality monitoring and feedback system



FAOSTAT is developing a standard data quality evaluation and monitoring processes of all statistical datasets currently found (or to be included) in the FAOSTAT system. A detailed description of the evaluation and monitoring process can be found in the paper presented at the Conference on Data Quality for International Organizations, Wiesbaden/Germany[4]. The standard data quality evaluation and monitoring processes covers both data submitted by countries and data that is 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 define 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 6).

A "quality stamp" will be attached to every dataset included in the FAOSTAT core. The quality stamp will inform the users that the data has met various checks to ensure quality. Documentation will identify how FAO procedures influence the overall quality of the data. FAOSTAT will also use innovative methods for visualizing and communicating data quality levels of various statistical data.

The 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 work being done by other international statistical offices on data quality monitoring.

Dissemination

FAOSTAT system products are being reviewed according to the FAOSTAT dissemination strategy focusing on: needs of FAO's information users; new functionality of the complete statistical system; new

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

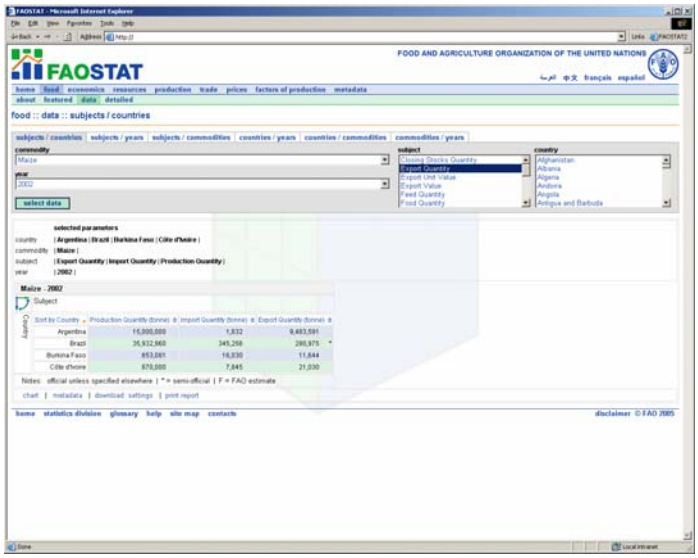
technologies and global trends in dissemination of information. Enhancing the functionality of the FAOSTAT portal will give users the possibility to perform more substantive analytical work. In particular users will find: An improved interface, with the facility to visualize FAOSTAT data or download data in different formats for further use in their preferred tools for analysis. Key country indicators in the form of precompiled reports will be a useful vehicle to assess a situation at a glance.

Users of the FAOSTAT portal will be able to select statistical data and metadata and subsequently to display and visualize the selected information using the function provided by the portal, or download the information and use it locally. For this reason, the portal will provide the following functions:

Data and metadata selection: Various data selection options will be implemented, from simple browsing of data and/or metadata to complex multidimensional selection options, including options to support flexible cross-domain querying.

Figure 7. FAOSTAT – New FAOSTAT Data Selection

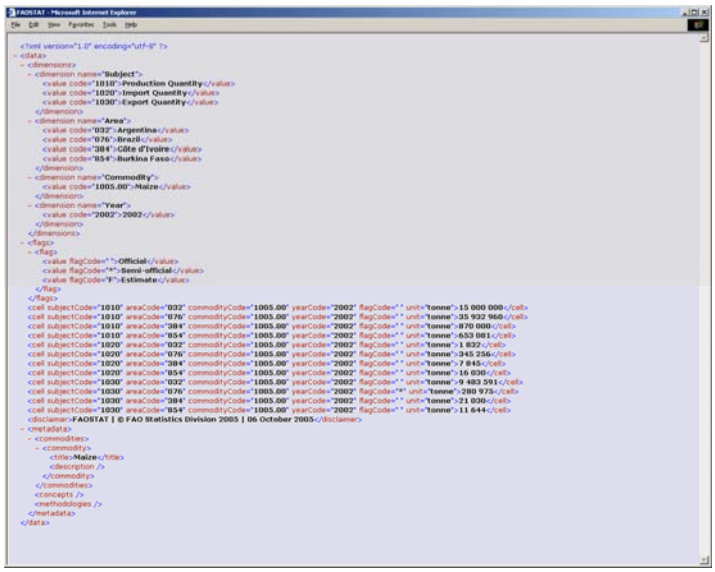
An example of cross-domain data selection is shown in Figure 8. It illustrates a few of the newly available functions such as sorting and charting.



Display and visualization: Functions to display metadata and data (including sorting, aggregating and filtering options) and to visualize selected data in different formats (tabular format, charts or maps) will be provided. Data should be displayed in a consistent style together with the metadata to ensure that information is presented to users in a transparent and comprehensible manner.

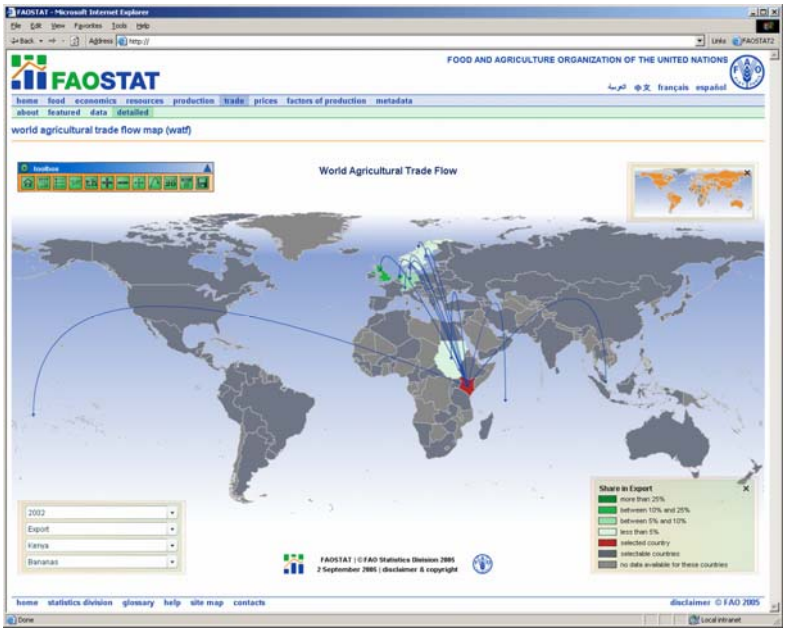
Download: Selected data and metadata will be easily downloadable in standard formats (CSV, Excel, XML, etc) in order that the user can easily save the information and use local tools to visualize and analyze it. Functionality and formats for reporting and printing (PDF) could be considered as well.

Figure 8. FAOSTAT XML output format



Specific dissemination products such the FAO Statistical Yearbook and the World Agricultural Trade Flow (see figure 9.) will be generated from the new FAOSTAT system and provide the statistical users in the FAO Statistics Division with the control over data series, countries time periods, metadata to be included in these products.

Figure 9. FAOSTAT World Agricultural Trade Flow

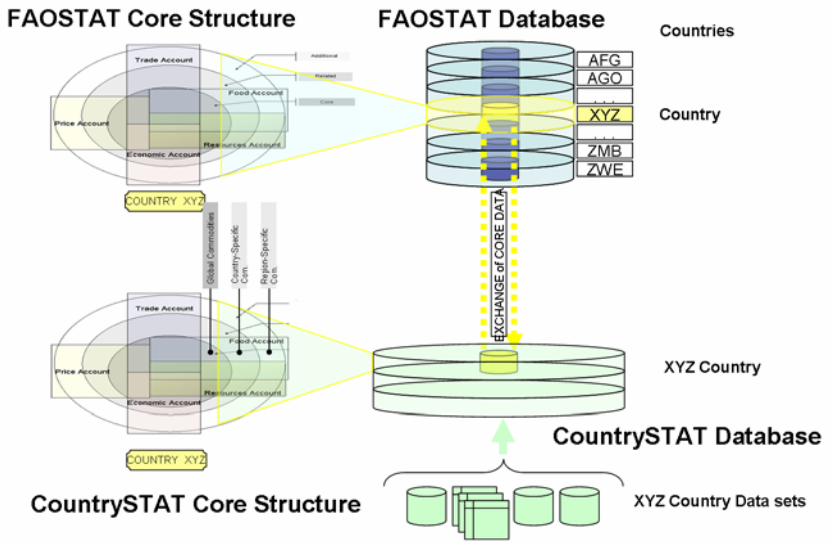


CountrySTAT System

The development of CountrySTAT, a scaled down version of FAOSTAT has provided a mechanism to integrate data, metadata and methodologies from various sources to assist national agricultural statistical systems for the purpose of facilitating data use by national policymakers and researchers. The CountrySTAT statistical framework is based on the FAOSTAT framework thus allowing interaction and data interchange (see figure 10.).

It aims to promote exchanges of data within countries, between countries and between FAO and countries. As with FAOSTAT, CountrySTAT will have a core statistical data series organised into statistical frameworks surrounded by statistical metadata. Countries will use tailored thematic modules as required for national policy or research. These modules will select data from various sources and combine them to analyze a particular theme or area.

Figure 10. FAOSTAT and CountrySTAT



CountrySTAT Pilot countries

CountrySTAT has produced prototype CountrySTAT modules for testing in a limited number of pilot countries. As part of the development process, CountrySTAT will provide conceptual and technical support for creating and implementing systems. Once CountrySTAT has been piloted, countries will be encouraged to implement the system. The CountrySTAT pilot studies are currently being run in the Kyrgyz Republic (National Statistical Committee), Kenya (Central Bureau of Statistics, and Ghana (Statistical Service). A report on these pilots will be prepared in December reviewing the experiences in each pilot country.

REFERENCES

Food and Agricultural Organization of the United Nations, *Modernization of FAOSTAT - An Update,* 2002. <http://www.fao.org/DOCREP/MEETING/004/Y6298e.htm>

Food and Agriculture Organization of the United Nations (2003), FAOSTAT Statistical Database,<http://apps.fao.org/>

Food and Agricultural Organization of the United Nations, *Presentations - Informal Expert Consultation on CountrySTAT*, 17 to 19 March 2003. <http://www.fao.org/es/ess/meetings/cy_stat_p.asp>

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

Food and Agricultural 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/acc