



UNITED NATIONS
DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS
STATISTICS DIVISION

ESA/STAT/AC.94/7
28 November 2003

**Meeting of the Expert Group on
International Economic and
Social Classifications
New York, 8-10 December 2003**

FAOSTAT2 Project and CountrySTAT

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

FAOSTAT2 Project and CountrySTAT

The FAOSTAT 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 UN, 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, leading to growing user frustration. 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 definition of requirements for a new FAOSTAT system commenced at that time. Interviews were conducted with representatives from all FAO organizational units that are currently providers of data to FAOSTAT as well as those who compile data that potentially could be included in a new system. Principal internal users of FAOSTAT data were also queried for their views on needed improvements to the system. A small representative sample of external users of the system was also contacted.

Objectives of the New FAOSTAT2 System

The anticipated end result of the FAOSTAT2 initiative is a new system that will meet FAO's Strategic Framework Objective of "*an integrated information resource base with current, relevant and reliable statistics, information and knowledge made accessible to all FAO clients.*" To achieve that goal, FAOSTAT2 must specifically address the following objectives:

- Provide an improved user interface, streamlined system processes, and a stable and reliable technical environment for the FAOSTAT Working System.
- Improve the quality of the data by providing robust tools for compiling, validating, estimating and analyzing data both at FAO Headquarters as well as at the country level.
- Incorporate new user requirements for system functionality and access to new data sets.
- Improve user access to FAOSTAT data by enhancing and creating new mechanisms for data dissemination, including access to data across domains.
- Enhance data integrity by ensuring that appropriate methodologies and data standards are consistently applied.

A Stable and Reliable Technical Architecture

The project approach recommends the use of components that have been created for similar projects already underway within FAO as a "quick start" method for building critical system functionality. Initial project outputs will address the more pressing technical problems and principal user requirements. Major improvements in system stability and reliability should be realized early in the project.

Incorporate New User Requirements

A number of new user requirements will be incorporated into the new FAOSTAT2. These include improved or new system functionality, as well as expanded data requirements.

Improved or New System Functionality: New functionality will focus primarily on tools to facilitate data capture and improve and monitor data quality: data exchange standards; flexible, yet comprehensive, data editing routines; data estimation support and tools to be used at country-level. Software will be developed to enable users to easily extract and analyze FAOSTAT2 data.

Expanded Data Requirements: The requirement analysis revealed a need for FAOSTAT2 to accommodate diverse types of data, which go beyond the standard cross section and annual time series currently maintained within the system. Many data sets currently maintained outside of the FAOSTAT system could be incorporated into an extended FAOSTAT data model. To accommodate these requirements, the FAOSTAT2 system should be capable of supporting the following:

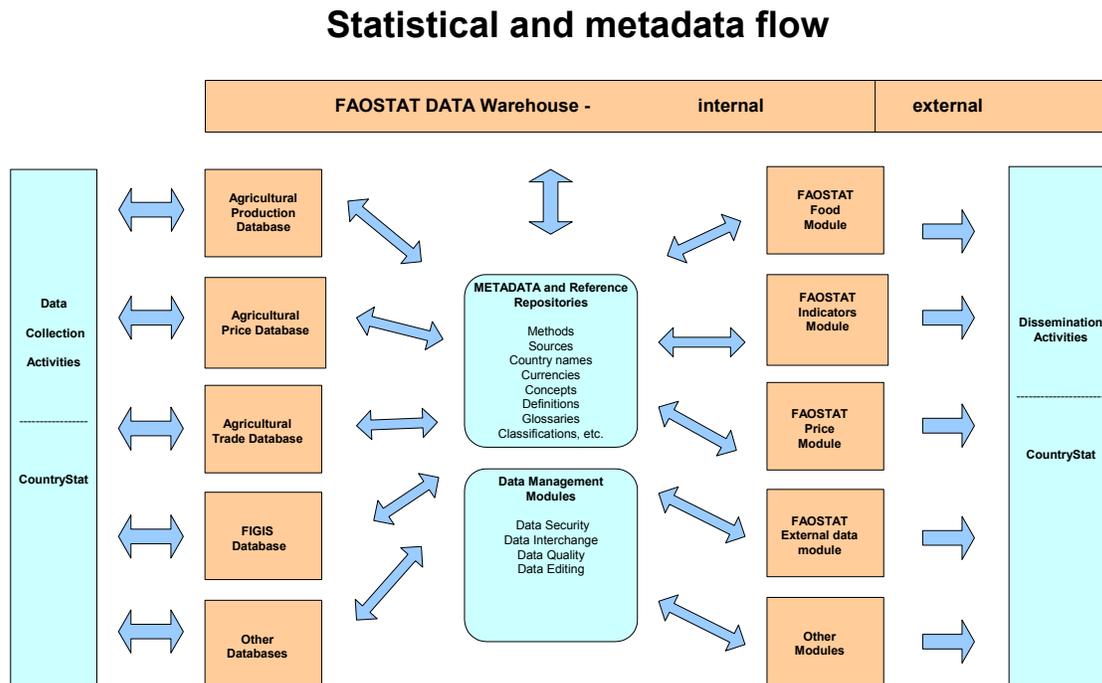
- *Data of varying periodicity:* FAOSTAT2 will support data capture at any frequency and interval, in addition to the standard annual time series data. This will include data captured monthly, on a market year basis, etc., as well as data captured on a less frequent basis.
- *Data projections:* FAO projections will be included for any category of data held in the system. This will provide a more complete picture of anticipated trends based on historical information.
- *Sub-national data:* 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.
- *Geo-referenced data:* FAO supports a number of spatial information systems that contain a wealth of information on geographic and climatic conditions impacting food and agriculture production. Integration of FAOSTAT statistical data with the various FAO spatial information systems will enhance FAO's analytical capacity in the food and agriculture arena. To support this requirement, the FAOSTAT2 system will provide the capability to attach a "geo-reference" to data maintained within the system.
- *Data from external sources:* Data available from external sources is often useful and necessary to assist FAO staff in their analyses. A flexible and easy-to-use mechanism is required to access data from such sources as the International Monetary Fund and World Bank and to merge them with FAOSTAT statistical data.
- *Meta data:* Meta data 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 FAOSTAT2 system.

- *Reference data*: Data used in supporting calculations and validations, e.g. conversion factors, country codes, editing rules, etc., will be available in and accessible to users of FAOSTAT2.

The FAOSTAT2 Data Flow is shown in Figure 1.

Figure 1. FAOSTAT2 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 via correspondence tables 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, such as the core food and price modules (see figures 3 and 4.).

Figure 2. FAOSTAT2 Core and Sample Draft Satellite Database System

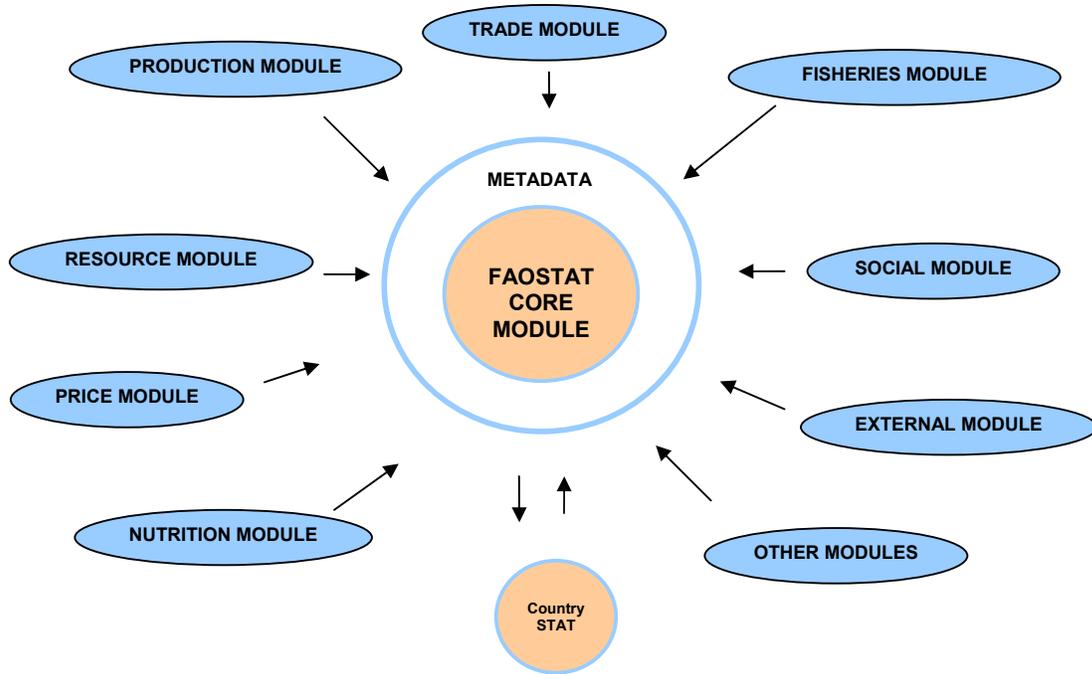


Figure 3. FAOSTAT2 Food Module

FAOSTAT FOOD MODULE

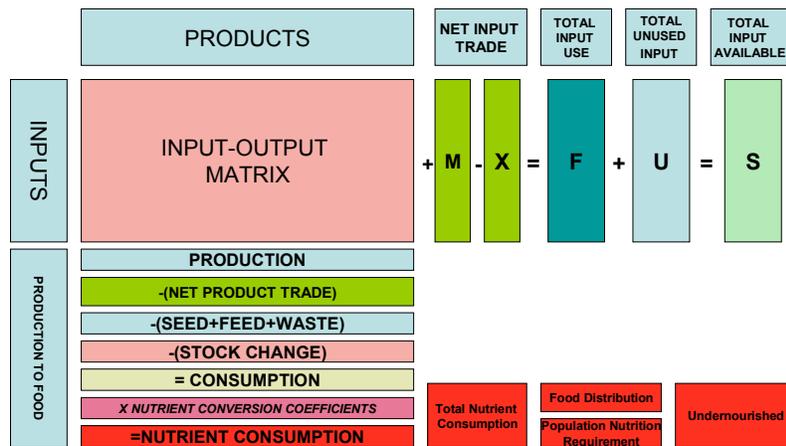
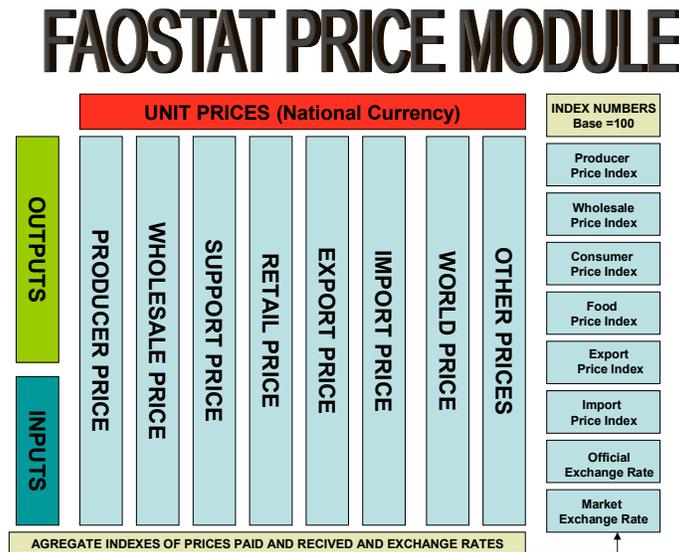


Figure 4. Example of a FAOSTAT2 Satellite Module



Improve Data Quality

Several mechanisms for ensuring and monitoring data quality (in terms of accuracy, completeness, etc.) are envisioned for the FAOSTAT2 system. All data entered into the system will be submitted to a series of rigorous editing and consistency checks. System editing will be performed in real-time to identify data errors early in the process and increase data entry efficiency.

Where a country has not provided complete times series data, the system will provide appropriate tools for estimating missing values. These tools will include interpolation and extrapolation algorithms as well as statistical models developed using analytical software packages.

The development and publication of an XML-based standard for the exchange of food and agricultural statistics will also facilitate improvements in the quality of FAOSTAT2 data. This standard will be distributed to member countries and other external partners who will be requested to provide their data in files formatted according to the published standard. Data provided in this fashion will be loaded, edited and validated, thus saving time and improving data quality.

FAOSTAT data quality ultimately depends on the quality of the data reported by member countries. This could be facilitated if countries entered and processed data online, which is currently possible, to a limited extent, through the Web-based virtual questionnaire. While this method will continue to be an option under the new FAOSTAT2 system, another option will also be made available: CountrySTAT, a scaled-down version of the FAOSTAT2 application, which will provide countries with functionality to compile,

validate, analyze and disseminate their national data. Outputs from CountrySTAT could then be easily loaded into FAOSTAT2 for further dissemination through FAO's web site, publications and CD-ROMs.

While countries will be encouraged to implement CountrySTAT, it is understood that all existing methods for data capture and transmittal - paper questionnaires, virtual questionnaire, and electronic files - will continue to exist under the new FAOSTAT2 system.

CountrySTAT

CountrySTAT¹ is a sub-project of the FAOSTAT2 project and will be developed in parallel to FAOSTAT2. CountrySTAT will be a scaled-down version of FAOSTAT2 that will be available to countries to implement the satellite modules as required. The major objectives for the CountrySTAT project are:

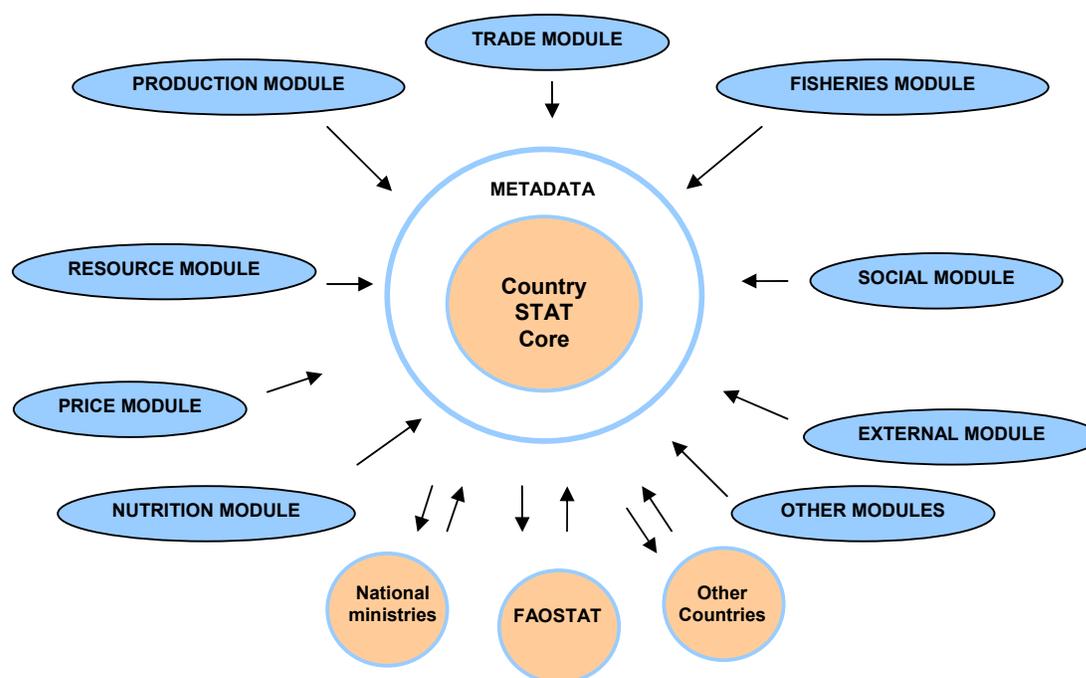
- Capacity building of member countries (coordination, harmonization and enhanced value to data);
- Two-way exchange of data - within countries, between countries and FAO, as well as between countries themselves;
- Facilitate data use by national policy-makers.

CountrySTAT will assist countries in developing a statistical information system containing available data and metadata relevant to agricultural policy, together with data from other extra-national sources, as well as FAO data relevant to the purpose.

As with the FAOSTAT2 project, CountrySTAT will have a core set of statistical data series similar to FAOSTAT2 set in a structured framework relevant to the countries specific agricultural situation (see figure 5). The core statistical data series will be surrounded by statistical metadata (classifications, definitions concepts, etc.) which will facilitate the flow of information both within countries, between countries, and between national statistics offices and international statistics offices. The various database satellite modules will be linked via the metadata to the core CountrySTAT. The satellite modules will be available for countries to add to the Core CountrySTAT as required.

¹ *Report of the Informal Expert Consultation on CountrySTAT*, 17 – 19 March 2003, Statistics Division, FAO, Rome, < http://www.fao.org/es/ess/meetings/fin_rep.asp>.

Figure 5. CountrySTAT Core and Sample Draft Satellite Database System



In terms of scope, CountrySTAT will provide a storage, verification, validation, analysis and dissemination system appropriate to a system of food and agricultural statistics at the national level. Thus, it will have the capacity to store sub-national and geo-referenced data. It will also have the basic tools that will be developed for FAOSTAT2 to verify, validate and derive the data (food balance sheets/supply utilization accounts) so that increased responsibility and ownership could be taken by the countries themselves for the quality and consistency of their data.

CountrySTAT will incorporate more detailed information than FAOSTAT2. It will contain sub-national data as well as other data such as credit, different level of prices, land rent, etc. and more breakdown of commodities. It will allow the coordination among various data collections at the national level so that FAO will not need to contact many data providers. FAOSTAT2 will be able to draw from a comprehensive database as the CountrySTAT product will contain more information.

CountrySTAT will be able to integrate within different IT environments of developing countries and to be easily updated, modified, maintained and sustained; taking into consideration what could be limited country resources and skills.

On the last issue, how to ensure consistency of data, by defining CountrySTAT as a country version of FAOSTAT2 this could only improve the consistency of data since the countries will be adopting an identical approach to data verification and validation as that used by FAO.

CountrySTAT is a sub-project under FAOSTAT2 and its development is seen primarily as a by-product of the development of FAOSTAT2. 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.

New Mechanisms for Data Dissemination

The original proposal for the modernization of FAOSTAT focused solely on the Working System, i.e., the underlying system used to compile, validate and analyze statistical data. The Working System clearly has the most pressing problems and is in need of replacement. However, the analysis of system requirements revealed the need to also address the redevelopment of the dissemination system.

The dissemination system is the visible portion of FAOSTAT, the component that most external users are familiar with. It is through the dissemination system that data can be accessed for further use. Enhancing the functionality of the dissemination system is a means to strengthen the capacity of internal and external users of statistical data to perform more substantive analytical work. Several opportunities exist to enhance and extend the functionality of the dissemination system. An improved interface, with the ability to visualize FAOSTAT2 data through graphs and maps, will enable users to highlight trends, anomalies and areas of concern. Key country indicators will also provide a useful vehicle to assess a situation at a glance.

Data dissemination mechanisms will be enhanced to support the automatic generation of high quality system outputs. New alternatives for publishing FAOSTAT2 data, such as dynamic online reporting and XML based data-interchange, will be introduced. Dissemination monitoring tools will be expanded to enable FAO to analyze patterns of system usage and to better tailor its services to meet the needs of its users. Initial work has started on the new internet user interface to FAOSTAT2. Examples of the new user interface are shown in Figures 6, 7 and 8 below.

Figure 6. FAOSTAT2 – Mock-up of New FAOSTAT Homepage.

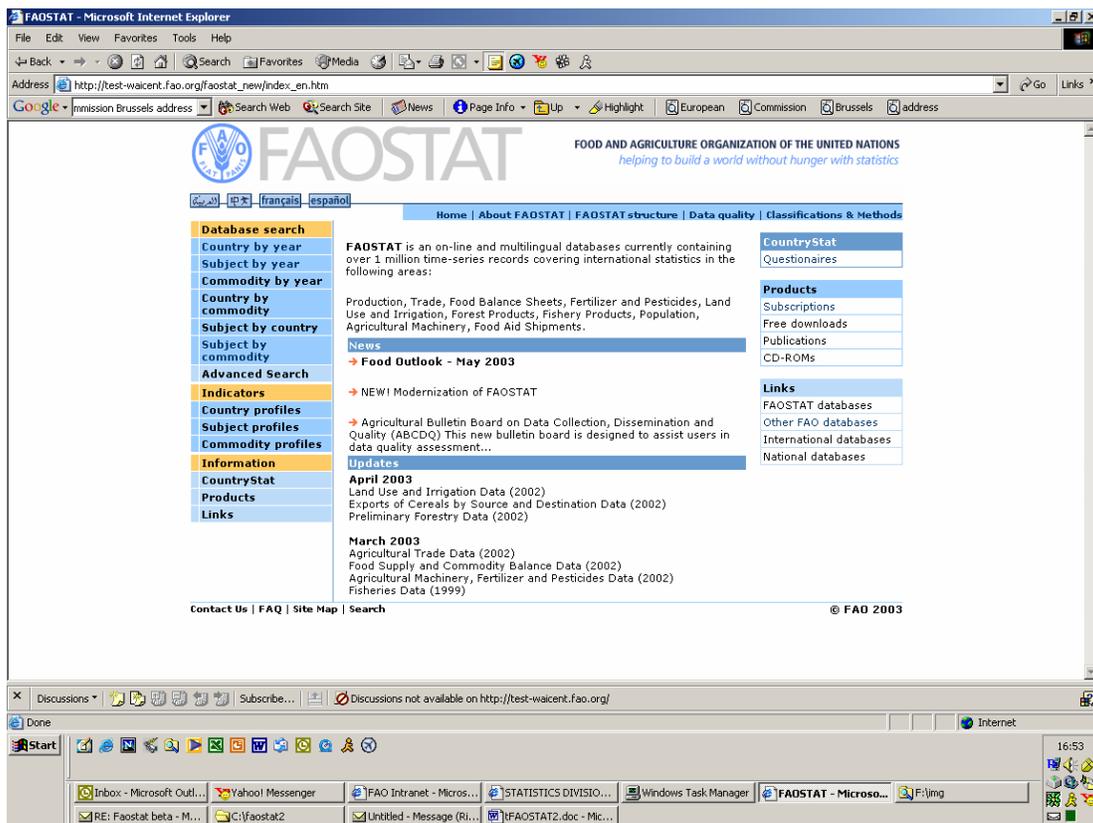


Figure 7. FAOSTAT2 - Mock-up of Data Selection of Artichoke Production by Country by Years.

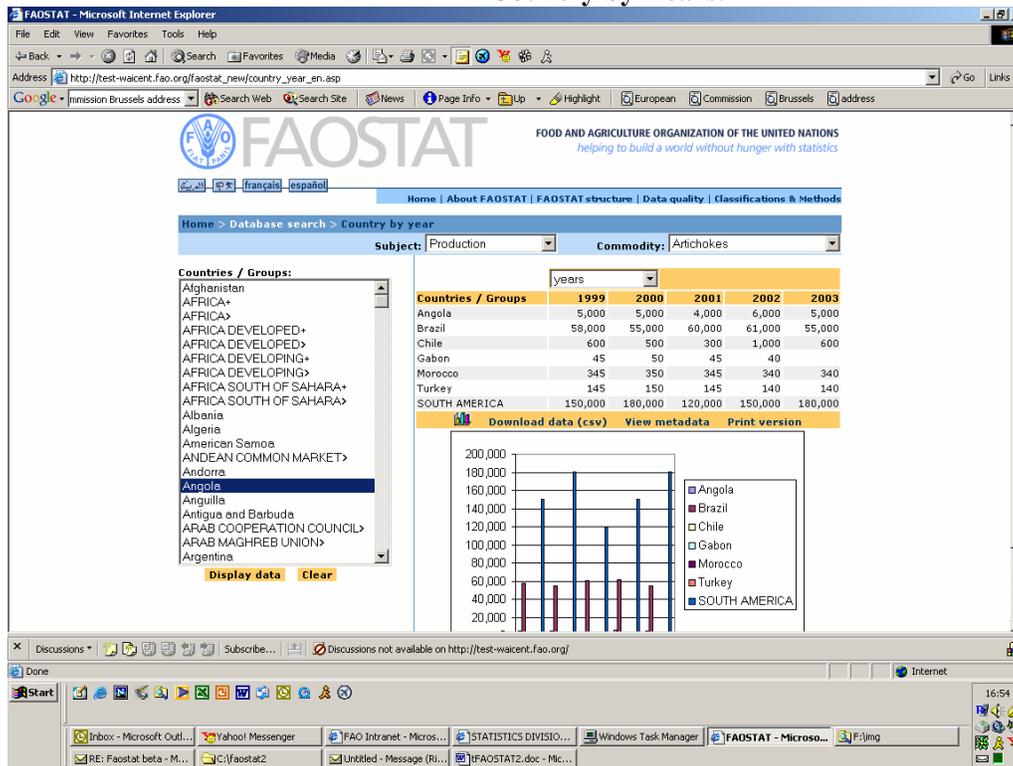
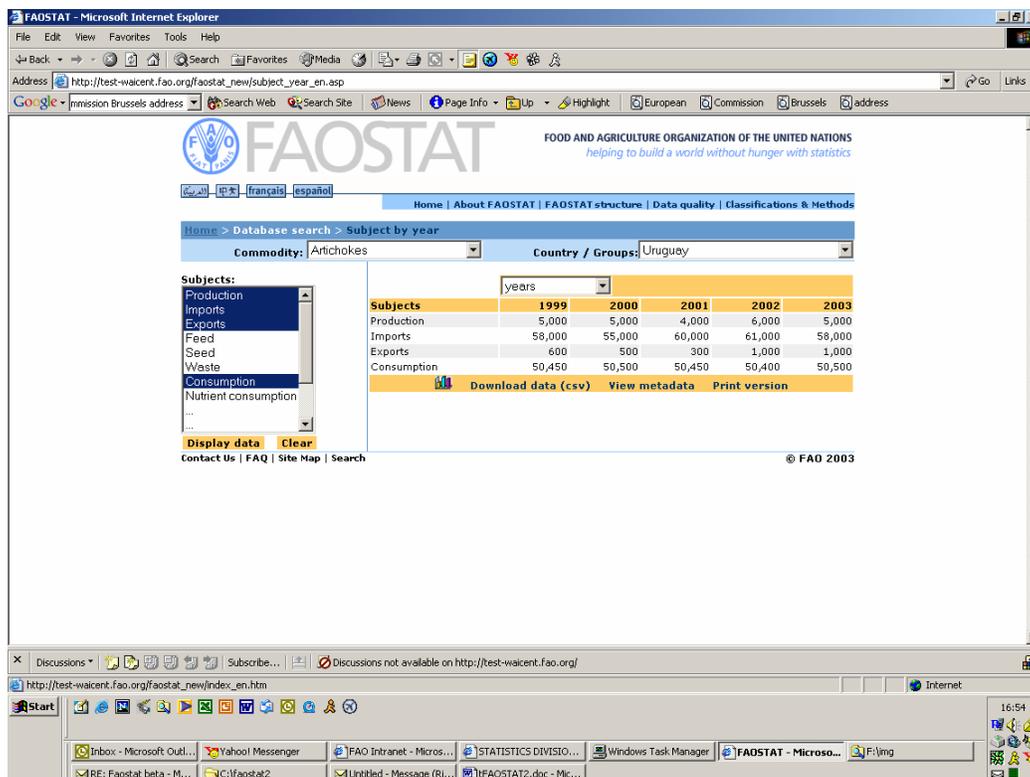


Figure 8. FAOSTAT2 - Mock-up of Data Selection of Production, Imports, Exports and Consumption of Artichokes for Uruguay for Selected Years.



Appropriate Methodologies and Data Standards

As the scope of the FAOSTAT system expands to include data from more diverse sources and to address the needs of more organizational units within FAO, the need to expand and reconfirm common data concepts, definitions, codes and methodologies becomes increasingly apparent. The FAOSTAT2 project offers an excellent opportunity to revisit existing methodologies and to identify areas for improvement, simplification and standardization. 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. Work has begun on reviewing current methodologies and classifications in view of future needs.

The Project Approach

FAOSTAT2 will be implemented using a phased approach. The project will be divided into sub-projects (statistical metadata, current Working System, dissemination system and integrated corporate database, CountrySTAT, etc.). Work on these sub-projects will proceed relatively independently of each other.

Initial outputs of the project will focus on improvements to the current system and resolution of the most pressing problems. The new system functionality will be capable of running on the current FAOSTAT system, thus enabling users to gain early experience with individual system components, before the complete, integrated system is delivered.

The FAOSTAT2 project will attempt to incorporate or use software components that have already been developed within the Organization. Rapid application development and prototyping will be employed, so that users can react to the proposed system and quickly provide feedback to developers.

The FAOSTAT2 development effort will be user-driven, involving key FAO staff at every level of the project. Data producers and data users will participate in the project from its inception, working as part of the project team to design and test the new system.

During the second half of 2003, work will concentrate on preparing the core framework and metadata for FAOSTAT2. Concurrently, methodological work will be undertaken on the Working System. It is expected that by early 2005, the FAOSTAT2 core and most of the satellite modules will be completed. During this development period, the CountrySTAT core and various CountrySTAT modules will be piloted in countries.