FAOSTAT and CountrySTAT: Integrated Global and National Food and Agriculture Statistical Databases

FAOSTAT e CountrySTAT: Archivi Integrati di Statistiche Agricole e Nutrizionali a Livello Globale e Nazionale

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

Riassunto: Questo rapporto riassume le principali attività intraprese nell'ambito dello sviluppo del nuovo sistema statistico FAOSTAT. Componente fra le più importanti del sistema informativo della FAO, FAOSTAT contribuisce al perseguimento degli obiettivi strategici dell'Organizzazione come la raccolta, "analisi", interpretazione e la disseminazione di informazioni riguardanti l'alimentazione, l'agricoltura e la nutrizione. Il rapporto presenta altresì la nuova versione nazionale di FAOSTAT, vale a dire CountrySTAT. Questi due sistemi informatici offrono innanzitutto informazioni e dati statistici relativi all'alimentazione e all'agricoltura a livello globale e nazionale. Essi forniscono inoltre una struttura che da un lato integra dati statistici relativi a vari argomenti tra i quali produzione, commercio, consumo, risorse e prezzi e dall'altro lato si propone di costruire un collegamento tra statistiche agricole e sistemi statistici generali e sistemi nazionali e globali di statistica.

Keywords: statistical database, food and agriculture, FAOSTAT, CountrySTAT, FAO

1. The FAOSTAT System

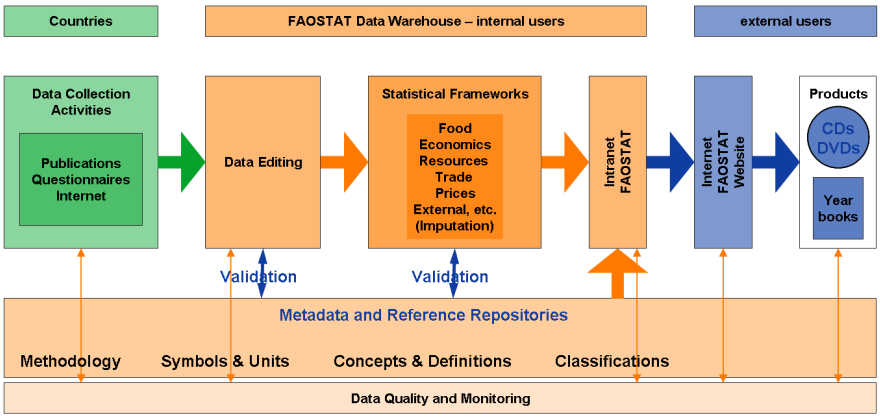
FAOSTAT is a well known product throughout the United Nations, statistical and academic worlds. Policy formulators, decision-makers and other stakeholders, both at the national and international levels are the primary users of the system. They regularly consult FAOSTAT as a data source for analysis and 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.

1.1. The Modernization of FAOSTAT

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 has started on the modernization of the FAOSTAT. The FAOSTAT project changed over time 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. The new FAOSTAT System provides updated tools for compiling, validating, estimating and analyzing data with an integrated centralized statistical metadata system. It is based on revised statistical methodologies which ensure data quality standards to be applied consistently. The system also provides an enhanced mechanism for improved user access to data.
An overview of the new FAOSTAT data flow (Figure 1) from national data to end users shows the scope of the modernization of the FAOSTAT statistical system.

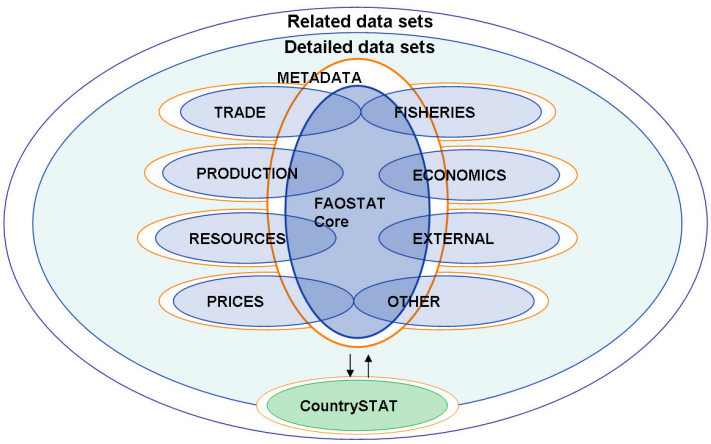
Figure 1: *Data Flow in the FAOSTAT Statistical System*



1.2. The FAOSTAT Family: The Core and the Satellites

The new FAOSTAT system revolves around a core FAOSTAT module (Figure 2.) with distributed satellite 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 contains standard statistical metadata elements to facilitate data interchange with the other database modules. All statistical data included in the core use the standard definitions and classifications and are fully integrated.

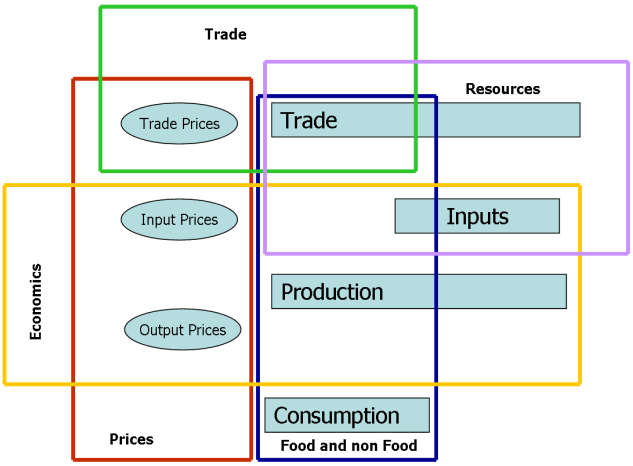
Figure 2: *FAOSTAT Core and Satellite Database System*



1.3. The Conceptual Framework of Integration

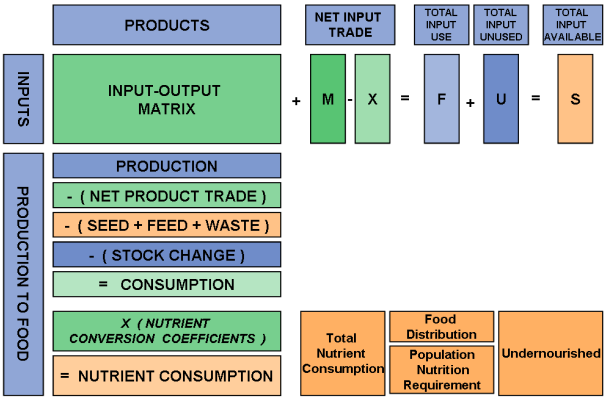
The integrated statistical frameworks center around major substantive domains of **Food** (Production, Utilization, Trade and Consumption); **Resources** (Input Availability, Utilization and Trade); **Economic** (Inputs, Input Prices, Production, Output Prices) **Prices** (Input Prices, Output Prices, Trade Prices), **Trade** (Imports, Exports). The integrated statistical frameworks constitute the core consolidated data in FAOSTAT (Figure 3).

Figure 3: *FAOSTAT Core Statistical Frameworks*



From these structured formats, key policy indicators are produced. The Core Food Module in Figure 4 provides an example of how this approach is implemented, producing estimates on undernourishment.

Figure 4: *Core Food Framework in FAOSTAT*



1.4. Statistical Methodologies

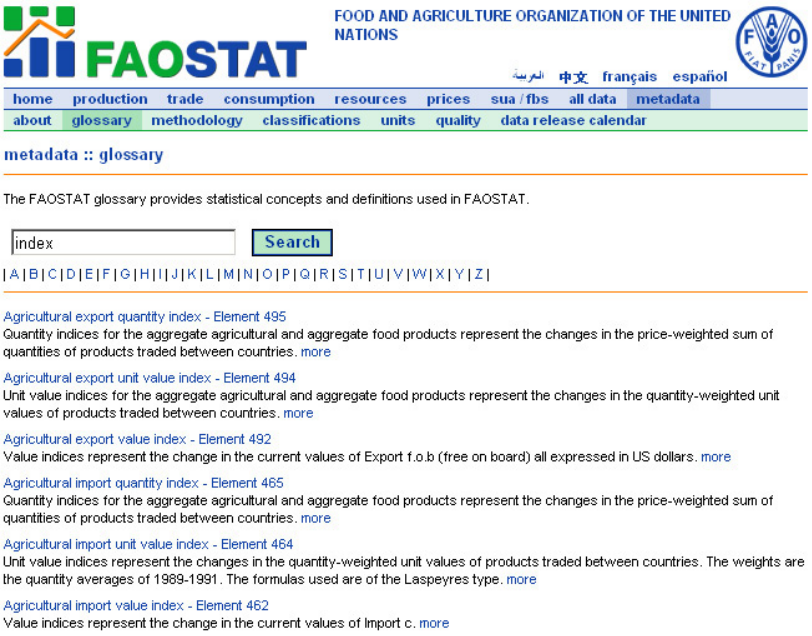
A peer reviewed, state of the art methodology framework for estimating, calibrating and balancing and thus producing a consistent coherent dataset within the statistical

framework domains has been developed, tested and implemented in collaboration with Bonn University. These statistical methodologies are used to generate datasets for the further generation of key statistical indicators in FAOSTAT. The approach is using robust estimation techniques adapted for situations where limited data is available. It is based on maximum entropy econometrics as well as Bayesian posterior density estimates.

1.5. Metadata

Structured, 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. FAOSTAT actively collaborates with the Statistical Data and Metadata Exchange (SDMX) initiative and has mapped terminology in the FAOSTAT Metadata system with the Metadata Common Vocabulary (MCV). The efforts draw on substantial work dedicated to building the FAO's ABCDQ system and on efforts in the international statistical metadata community (Figure 5).

Figure 5: *FAOSTAT Metadata Glossary*



1.6. Classification

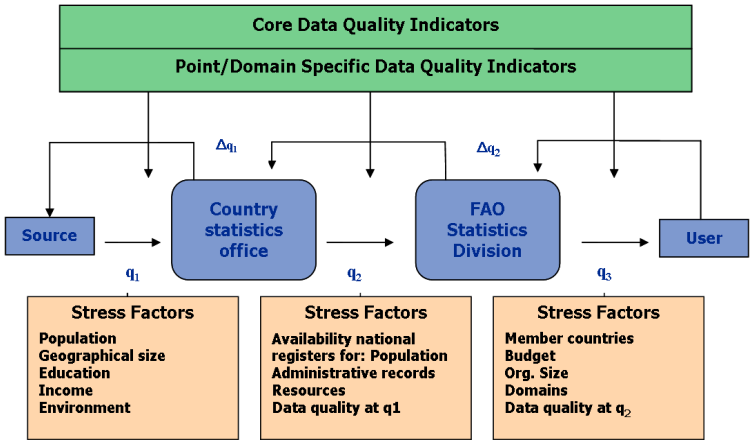
The FAOSTAT list of commodities is a list of agricultural products that serves the specific needs of agricultural production statistics. The list was inspired by SITC and adapted by FAO where needed; in most cases additional detail was added to the SITC

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

1.7. Data Quality

FAOSTAT is developing a system which aims to monitor data quality in all stages and aspects of the international data process (from Country data 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 (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 System*

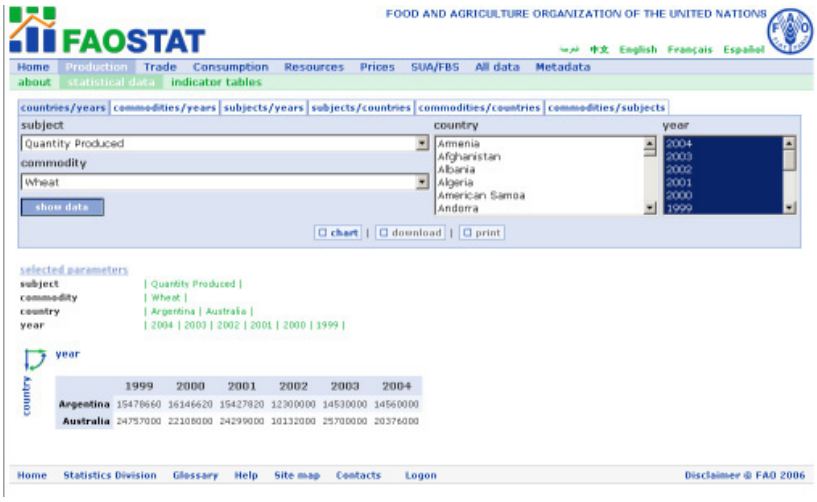


1.8. Dissemination

FAOSTAT system products are reviewed according to the FAOSTAT dissemination strategy focusing on: needs of FAO's information users; new functionality of the complete statistical system; new technologies and global trends in dissemination of information. Enhancing the functionality of the FAOSTAT portal gives users the possibility to perform more substantive analytical work. In particular users will find an

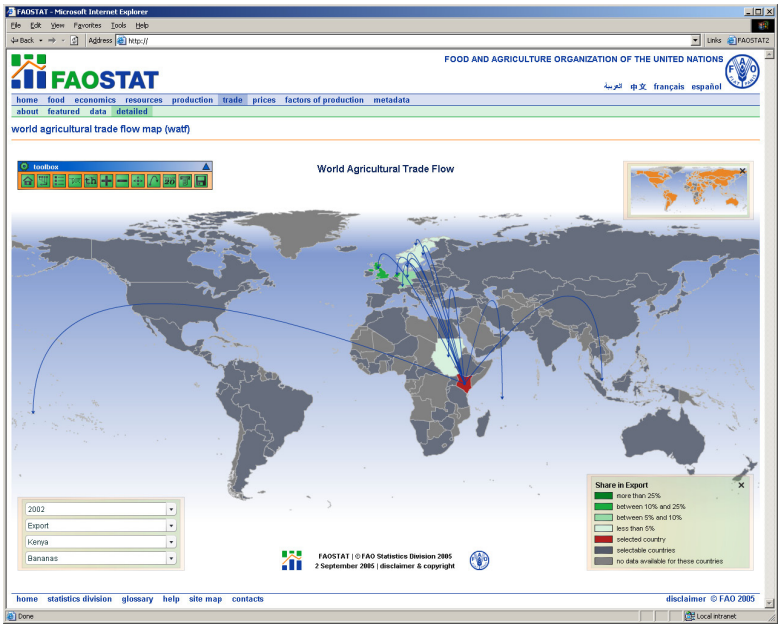
improved interface, with the facility to visualize FAOSTAT data or download data in different formats for further use in their preferred tools for analysis (Figure 7).

Figure 7: *New FAOSTAT Portal*



Users of the FAOSTAT portal are now 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. 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) are provided. (Figure 8).

Figure 8: *FAOSTAT World Agricultural Trade Flows*



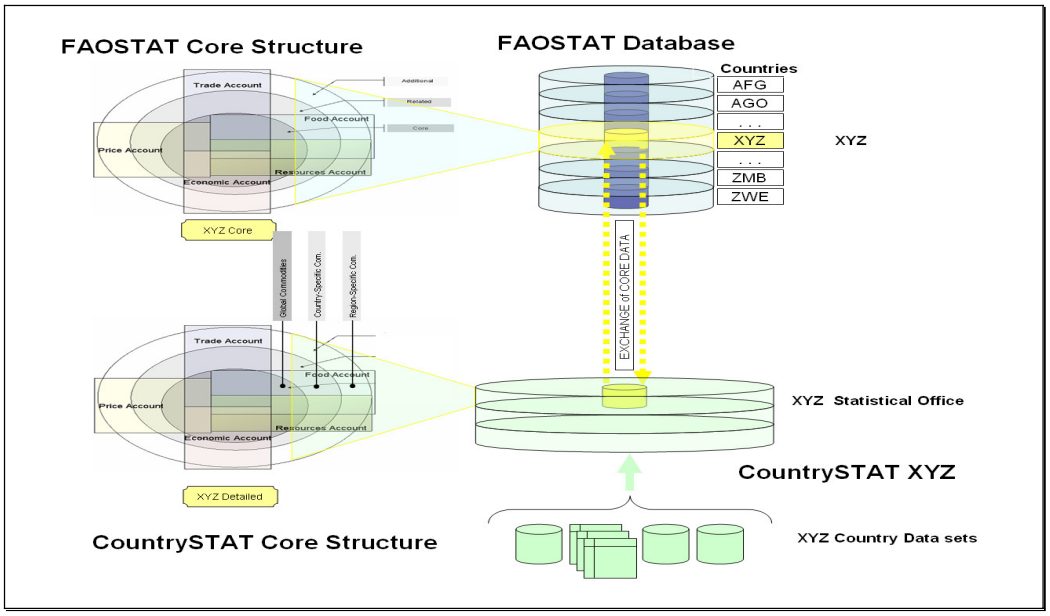
2. The CountrySTAT System

CountrySTAT is an integrated national food and agriculture statistical information system ensuring harmonization of national data and metadata collections for analysis and policy making. The web based system uses PX-Web has been tested in the statistical offices of Kyrgyz Republic, Kenya, and Ghana during 2005.
CountrySTAT provides the capacity to store and report on country data that is captured at the sub-national or administrative unit levels. CountrySTAT was developed as a sub-project under FAOSTAT and its development was seen primarily as a by-product of FAOSTAT. By definition, the two entities FAOSTAT and CountrySTAT evolve together in complete synchronization of their principles on data and metadata classification and organization.

2.1. The Methodological Framework

The organization of the data collections in the CountrySTAT system corresponds to the analytical framework of core and satellite modules which are commonly shared by CountrySTAT and FAOSTAT approaches (Figure 1) and builds the semantic link between FAOSTAT and CountrySTAT data and metadata collections (Figure 9). The core framework finds its manifestation in the webbased graphical user interface of the CountrySTAT System under section "CountrySTAT Core".

Figure 9: *CountrySTAT and FAOSTAT*



In CountrySTAT Core national level, statistical matrixes are organized under logical themes and their sub-themes. It is important to note that in all CountrySTAT participating offices, each CountrySTAT System is structured in the same way, thus once a user has familiarized himself with CountrySTAT Core Structure he will be able to navigate immediately in any other CountrySTAT system and also be familiar with the Core Structure of FAOSTAT. Further to the Core, countries are provided with the

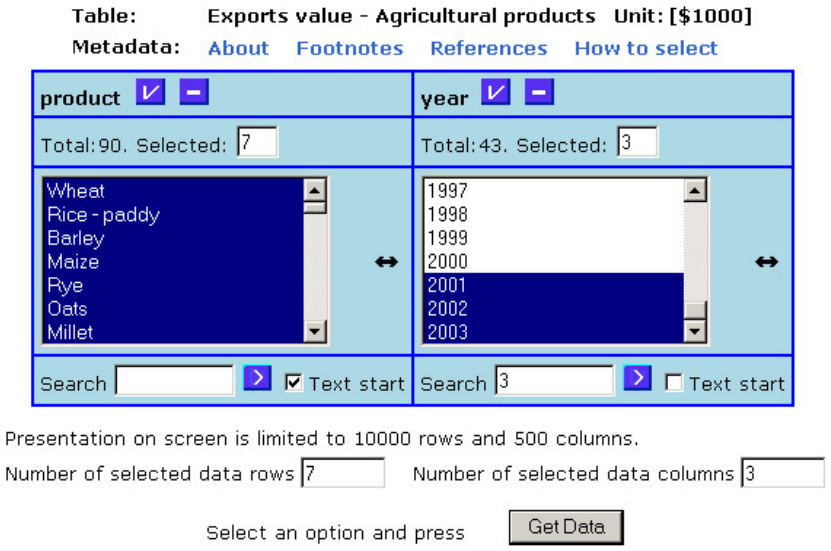
flexibility to organize their matrixes with extended levels of statistical data in their way of looking at the subjects under "CountrySTAT – Details". These satellite modules with detailed, specialized data collections are visualized correspondingly to the Core into folders that can be customized to requirements of the country. It is worth to mention that each matrix can be associated to more than one thematic folder, within as well as across the core and detailed thematic modules.

2.2. The Operational Features of CountrySTAT

2.2.1. Dynamic Generation of Matrixes

The user can choose the relevant matrix from a thematic folder. A **query dialogue window** opens (Figure 10) where the user can select within each variable the relevant values and execute a query to the database, which is actually a human readable structured ASCII file. The queries result are handled by PX-Web the underlying engine of CountrySTAT System and send back via the interface as **tabulated data sets** which can be rendered according to the analytic needs of the user using layout templates, structured and free transposition of variable vectors.

Figure 10: *Interactive Query in CountrySTAT*



The system also offers **quasi-fix tables** where the user select among a ready ideally tables by predefined the system administrators. These **summary tables** (Figure 11) offer the user to jump to the fully interactive query dialogue window (Figure 10) allowing the users to perform specialized investigations from the matrix. An important feature is the same table can be associated to more than one folder.

Figure 11: *Predefined Summary Tables in CountrySTAT*

Dig deeper

Exports value - Agricultural products by product and year

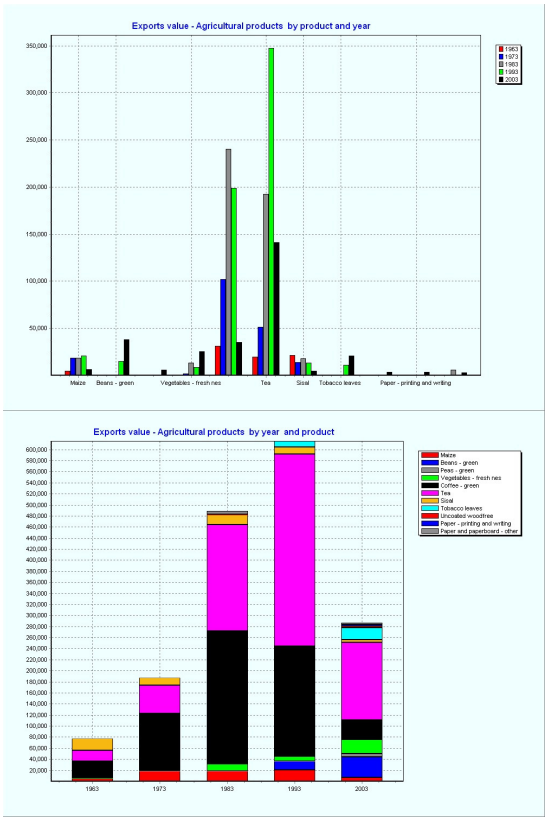
	1963	1973	1983	1993	2003
Tea	19,856	51,129	192,587	347,432	140,931
Beans - green	0	0	0	15,073	38,207
Coffee - green	31,285	101,985	240,170	198,847	35,076
Vegetables - fresh nes	315	1,997	13,277	8,558	25,526
Tobacco leaves	321	31	1	10,753	20,828
Maize	4,760	18,224	18,237	20,866	6,397
Peas - green	0	0	0	717	5,584
Sisal	21,096	13,671	18,143	13,066	4,771
Paper - printing and writing	0	174	723	92	3,283
Uncoated woodfree	0	0	0	0	3,178
Paper and paperboard - other	13	214	5,811	21	2,819

Footnote:
Food and Agriculture Organization of the United Nations
Project CountrySTAT
Viale delle Terme di Caracalla 00100 Rome, Italy
contact: Kafkas Caprazli
e-mail: kafkas.caprazli@fao.org
phone: +39 06570 54916

2.2.2. Analytical Charts

The resulting tables can be subject to displayed using the build in charting function. The charting options include bar, horizontal bar, lines, areas, points, pyramids, radar, histogram, stacked bars, horizontal stacked bars. (See below Figure 12). Each chart can be resized and reused in regular office software.

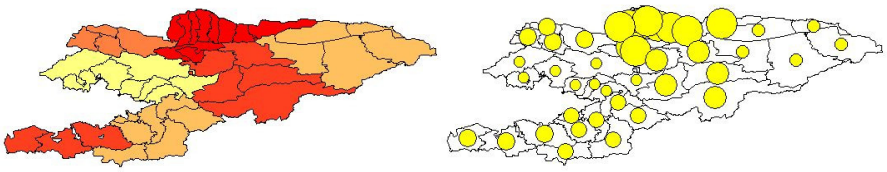
Figure 12: *Chart Samples from CountrySTAT*



2.2.3. Analytical Maps

In cases where data have been associated with geographically distinct areas like regions, provinces, districts of a country, the resulting tables can be displayed on scaleable maps (based on ESRI shape files). These options work for the range of single layered map visualizations: Quantile as well as equal interval corophleth maps and circle symbol maps. (Figure 13). Each map can be resized and reused in regular office software.

Figure 13: *Map Samples from CountrySTAT*



2.2.4. Data and Metadata Exchange

The resulting (tailored) tables are generated in HTML format; however, all can be changed into common as well as specific exchange formats, i.e. MS Excel, CSV, and PC-Axis formats. The tables that are downloaded contain the exact selected vectors and their ordination, which is a big help for restructuring large scaled multi-dimensional tables, especially if a statistical office decides to explore the maximum range of 16 variables per matrix. All exchange formats contain the full information on data and metadata. The **underlaying (entire) matrix** can be harvested entirely by using XML/SDMX format which is a static file prepared by system operators when data are updated by the underlaying engine PX-Edit applying an SDMX variant. SDMX is a XML file validated against the DTD (data type definition) as defined by the Statistical Data and Metadata Exchange initiative.

2.2.5. Integrated and Referenced Metadata

CountrySTAT System utilizes the PC-Axis format using many-dimensional matrices, also called cubicles or boxes Metadata are integral part of the matrixes: Values of the commodity variables utilize the classification and coding system as applied by FAOSTAT.Thus labeling variable values in local languages as well as aggregations can be operated safely. Values of the regional variables utilize the GAUL classification, where this is not applicable yet the national classifications are utilized consistent with those in FAOSTAT.

2.2.6. System Scalability

The CountrySTAT project tested scalability in two levels: Scalability of a single matrix: A table simulating trade data of 184 countries over 40 years for 120 commodities (883 200 records, each with 8 digits) was generated and handled within parts of a second on a laptop. The performance on a real server should be expected to be faster. The limiting factor in the display was the generation of a table structure by the internet browser. Scalability of matrix structures: 2 000 full-scaled tables with simulated data were generated and placed into different folders. The navigation structure through and across these matrixes was created within less then a minute by the systems operator.

2.3. Developing Partnerships for CountrySTAT

The CountrySTAT statistical system aims to match the needs of policy makers and researchers at the national level based on the FAOSTAT statistical methodologies. FAO has set up prototypes of CountrySTAT systems in partnership with national statistical offices of Kyrgyz Republic, Kenya, and Ghana. These pilot studies are representing typical conditions of countries in which the primary sector is of major importance for the food security as well as the development of the country. The three CountrySTAT pilot projects in Kyrgyz Republic, Kenya, and Ghana indicate that the statistical framework and the applied technology as well as the outreach approach are suitable to meet the needs of FAOSTAT and the needs of the Statistical Offices. Data and metadata can be handled easily, classification and codes can be accommodated in the low IT demanding system, and sustainability can be expected. The system is now ready for wider implementation subject to the availability of funds and interest from countries.

References

Food and Agricultural Organization of the United Nations (2002), Modernization of FAOSTAT - An Update <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 (2003), Presentations - Informal Expert Consultation on CountrySTAT, 17 to 19 March <http://www.fao.org/es/ess/meetings/cy_stat_p.asp>

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

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

Food and Agricultural Organization of the United Nations (2004), 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 <http://unstats.un.org/unsd/acc>

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

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

Golan A. Judge, G. and Miller, D. (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley & Sons Inc.

Heckelei, T., Mittelhammer, R. Britz, W. (2005) *A Bayesian Alternative to Generalized Cross Entropy*, Bonn

Heckelei, T., Mittelhammer, R., Britz W. (2005) Solutions for Underdetermined Models, Paper presented at the 89th EAAE Seminar, February 3-5, Parma, Italy

Hodrick, F.J. & E.E. Prescott (1997) "Postwar U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit and Banking*, 29, 1-16