WP.1
26 April 2004

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (i): Web technology in statistical information systems

ConIstat-on-line (http://con.istat.it): Short term statistics time-series data base in WEB environment

Submitted by the National Institute of Statistics (ISTAT), Italy [1]

1. INTRODUCTION

The timely availability of short-term economic data is one of the main problems in the field of diffusion of high frequency data. Moreover a quick and easy access to the data is a strong requirement by the users. For several years, ISTAT (Italian National Institute of Statistics) has disseminated time series by IBM mainframe as a dialogue application using terminal or personal computers in terminal emulation. In 1997 ISTAT decided short-term time series dissemination on "Conistat" CD-ROM database. This data base was implemented by a two-tier client-server architecture operating under Windows.

The user had to install in his own computer both the database and the applications that enable the data search. Based on the calendar of press releases, it was possible to download the updates from the bbs.istat.it site on the Internet. However, this operation took a lot of time because the user was forced to update the entire data base even if he was only interested in certain time-series.

The development of the WEB technology, as well as the availability of increasingly more sophisticated, efficient and cost-effective Web development tools and high-performance and low-cost hardware, has led ISTAT to redesign this data base in a Web environment.

During planning phase, it was noted that several functions already present in the CD-ROM version have been extensively appreciated by users. Therefore, within the framework of Web technology, these functions have been maintained: *the user interface and the server-side functions have been designed carefully to reach the right compromise between what the CD-ROM version currently offered and what the on line data base could offer*.

About the system design in a Web environment, particular consideration was given to several characteristics that are not very important in stand-alone and off-line environments and that were thus not considered in ConIstat design for the CD-ROM version, such as ***user management***, ***management of concurrency*** and the ***use of a stateless protocol*** such as HTTP.

[1] Prepared by Francesco Rizzo (rizzo@istat.it)

The basic problem concerning the entire project has implied adapting a technology created to use hypertext documents to seek structured information in a databank.

The solution was to organize the research within the database through a ***search tree*** with only a few branches, so that the user could reach the desired information after visiting only a few pages.

A very convenient way of conceiving the application was to divide it into the different functions that could be planned to perform: many client-server applications are based on the ***transactional processing model***. This makes it possible to divide the functions into discrete transactions and manage them from the start to the end.

In Web technology, it is very helpful to consider the various pages as a single group of transactions. The work unit performed by each page can be considered a single transaction, even if a request for information or user authentication is involved.

The design of the system considered:
- functional features;
- technological features.

This design phase was backed by the creation of various prototypes that ascertained the feasibility of the system and its usability.

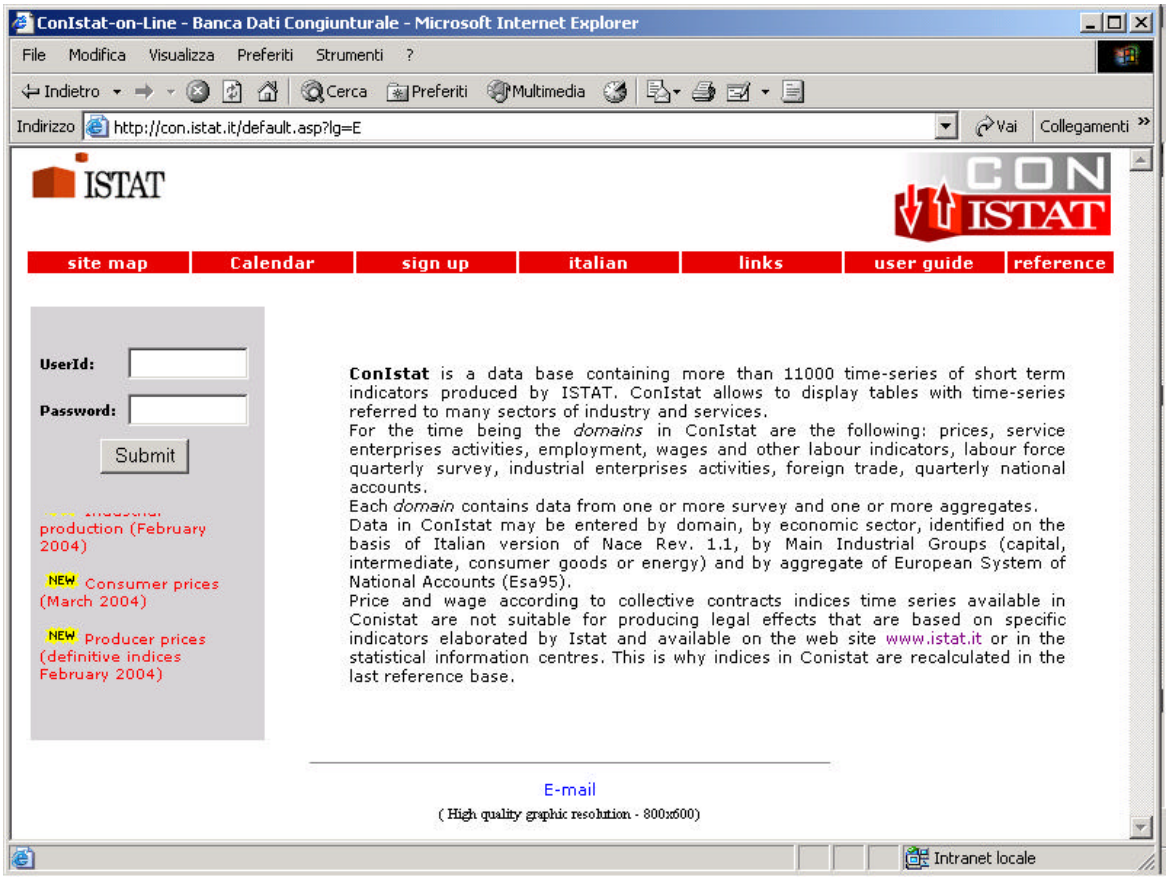
2. FUNCTIONAL FEATURES

The main functions that a user can perform with the navigator are:
- Personal authentication by inserting a user ID and a password, or registration by filling in an on-line form; in this case, the proper user ID and password will be sent immediately to the new user by e-mail;
- Choose of the right language (Italian or English)
- Data search by the ***Ateco*** (Italian version of NACE) code, by ***Main Economic Industrial Groupings (MIGS)***, by ***Esa95*** or by ***aggregation dominion***;
- Time-series displaying and comparing from different Ateco, Migs, Esa95 or different aggregates;
- Data selection by choosing the time-series and the appropriate time interval (starting period – starting year, ending period – ending year);
- Simultaneous management of time-series with different frequencies;
- Full metadata visibility;
- Management of both raw and seasonally-adjusted time-series;
- Time-series download in text (.txt) and csv file format;
- Put extracted data on a graph;
- Saving of one or more search paths in order to recall them during the following work sessions;
- Management of both short term indicators with base 1995 or base 2000.

All the above-listed functions have been implemented through a search system (navigation) designed to be logical and intuitive, also allowing beginners to use the system. The Conistat homepage, specific for advanced internet services, presents, on the top-bar of the layout, various information and service hyperlinks to other site-pages as user registration form or metadata description page, or the calendar or the user-guide. There's also an English homepage version, reachable from the Conistat homepage by a hyperlink on the top-side bar.

In the centre of the homepage there's the login interface in which the user fills his user-Id and his password, transmitted by e-mail after the user registration.

The down-side of homepage layout contains a short overview on the data bank and the webmaster e-mail hyperlink.



Whenever a user requests the home page of the data base he opens a work session. Through the session, the user can perform some operations, such as defining some queries that he can recall in a following session.

The sessions are characterised by the following parameters:

- Time-out: if the user does not interact with the data base for a time period longer than the one indicated in the time-out, the session is closed automatically. This parameter is set at the value of 5 minutes.
- Maximum number of selectable series. This parameter, set at twenty, specifies the number of series for whose data can be extracted simultaneously. Once the user has placed the various series in the shopping basket, he can extract the data for only five series at a time. . In the next months the number of series that can be viewed and downloaded will be strongly expanded.
- Number of searches that can be stored. After placing a given number of series in the shopping basket, the user can save them as a single search. Up to twenty searches can be saved. This operation allows the user to recall the saved searches in a future session. As a result, the user is not forced to navigate inside the databank to reach the desired data.

3. SYSTEM ARCHITECTURE

ConIstat is a complex Information System that makes available to users a data base containing more the 9.000 time-series of short term indicators produced by ISTAT.
The system includes an external component (internet) and an internal component(intranet).
The first one involves ISTAT external users and it's characterised by interrogation functions that allow users to get information requirements through a web-browser tool.

The intranet component has got the same data-search functions of the external one, but it runs in a different environment compatible with specific functions performed in order to manipulate the data to be included in the database.

3.1 External component

This part of the system's architecture includes:

- A database for storing data.
- A software program to perform user queries.
- A software program to perform updating operations.
- A software program to perform database dissemination activities and to collect the site statistic.

3.2 Internal component

This part of system has got two functionality modes:
a) Central functions performed on a machine work, as a server
b) Local functions distributed through the person in charge of Service Units involved in the production of statistical information to be stored in Conistat.

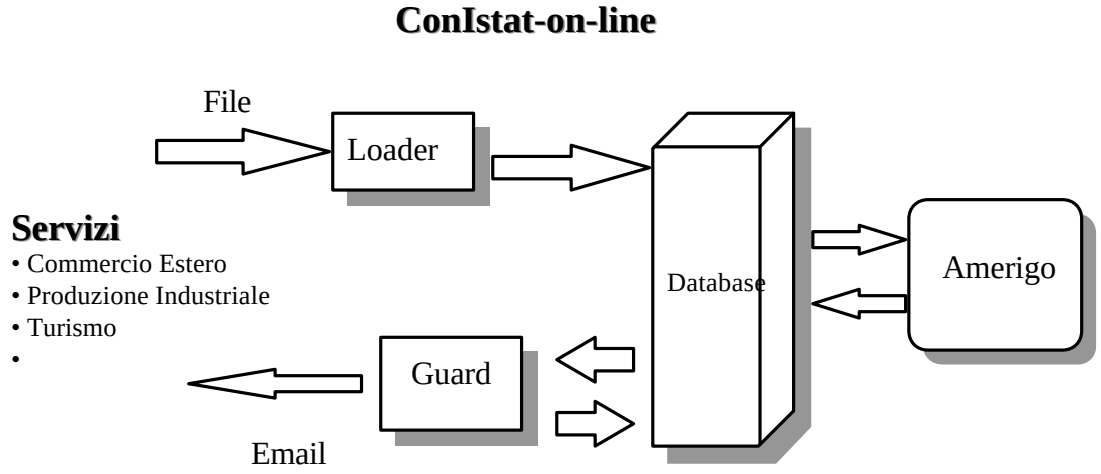
Each responsible uses a client-server program (Check), that performs a quality control of the statistical information and arrange the data in a database compatible format. Moreover there's a pilot software (Data sender) that builds time-series sets in ASCII format to be periodically sent to users that require them. This program runs in full-unattended mode and does not require interactive operations.

3.3 Technological features

The system is composed of several applications and a DBMS. The Department employees in charge to send data to Conistat, prepare the files according to well-defined specifications. These files are deposited in the server in which the database resides. On a daily basis, the "***Guard***" program interrogates a database table in which the calendar of "press releases" is stored. When it finds dates in the press release that have already elapsed without data-loading operation, an e-mail message is automatically sent to the managers of the Department that should already have sent the files by those dates.

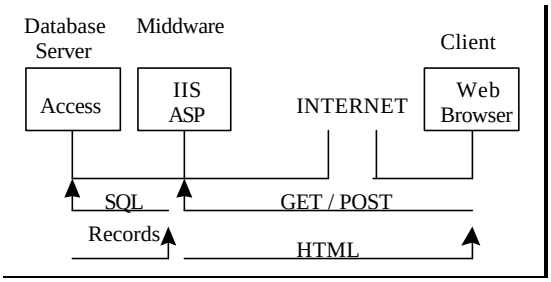
With each press release, an operator manually launches the "***Loader***" program that loads the new data in the database (an automatic procedure will be implemented in the future).

With any browser, the user can connect himself to the databank using "***Amerigo***", a software program that allows navigation among time-series.



The system is based on a three-tier client-server type operating under a Web environment structured as follows:

- ***client***: a Web browser;
- ***middleware***: a Web server and an application server;
- ***server***: a Database Management System



Through the Post or Get method of the HTTP protocol, the client sends requests to the Web server.
The Web server analyses the requests and through the application server, it interprets the scripts that are called up by the browser.

If the scripts contain information requests to the database, SQL queries are launched to the database.
The database returns the requested records to the application server.
The Web server formats the results in HTML and sends them to the client which, through the browser, makes them available to the user in a legible form.

The system has been designed to operate with all the databases that are interfaced through ODBC drives. Therefore, if a decision is made in the future to convert the database to run under Oracle, the changes to be made to the system will be marginal and will consist of:

- suitable configuration of the SQLNet and of the ODBC drive;
- set-up of the appropriate connection string (currently positioned in the Global.asa file);
- revision of some of the SQL queries.

3.4 Software/hardware features

The Web server is Microsoft's Internet Information Server 5.0 running under Windows 2000 server. The application server is Active Server Pages (ASP). The application is written to be as browser-independent as possible and must run under both Netscape 3.0 and MS Internet Explorer 3.02 version or higher. The browser must support certain HTML constructs such tables and forms, and if it is JavaScript-enabled it can optimise the interaction with the web server. The browser must also accept cookies that are sent back to the original server as a fundamental requirement of MS ASP.

The database uses the SQL JET 4.0 search engine of MS Access 2000. In any event, the system has been designed to operate with all ODBC-compliant databases. In particular, the system is already set up to work with DBMS Oracle. To optimize costs, both the database server and the middleware operate on the same computer: a two Pentium IV Pro at 900 MHz with 2 GB of RAM.

3.5 Design elements of the Web application

The system has been developed with a distinction between the server-side functions (Visual Basic Script in ASP environment) and the client-side functions (JavaScript):

- Server-side
- ✓ Interrogation functions.

  To perform these functions, objects from the ADO family were used (ActiveX Data Object) as well as the ODBC drive in pooling mode (to optimize the connection with the database).

- ✓ Functions for presentation of the requested data.
  These functions were developed using the HTML tag <Table> … </Table>. The selected time-series are presented in a table form. In the same table, the system can present time-series with different frequencies.
- ✓ Functions to manage the user session.
  The system can store the variables. In addition, an algorithm has been developed to take the time-series selected by the user as he navigates and place them in a container, maintaining them for the entire session.
- ▪ Client-side
- ✓ Functions for validating the operations performed by the user.
  These functions have been implemented in client-side JavaScript. However, to guarantee compatibility with browsers that do not support JavaScript, the same functions have also been developed in server-side Visual Basic Script.

4. INDICATORS AVAILABLE IN CONISTAT

Data available in Conistat are times series concerning the main economic areas analysed by Istat surveys in the short term indicators field. About the domains, the type of data, the available breakdown, the times series number and length and the data adjustment, it's possible to summarise the following framework (until now the series in Conistat are 9574):

Prices
1.1 Consumer prices (298 series)
1.1.1 Consumer prices for whole nations
1.1.2 Consumer prices for blue and white collar workers households
1.2 *Producer prices (163 series)*

Services
2.1 Retail sales trade index (405 series)
2.2 Other services (5 series)
2.3 Tourism (45 series)

Employment, wages and other labour indicators
3.1 Large firm labour indicators (3320 series)
3.2 Wages according to collective contracts (251 series)

Industry
4.1 Industrial production (384 series)
4.2 Industrial turnover (432 series)
4.3 Industrial new orders (258 series)
4.4 Industrial orders stock (255 series)

Foreign Trade *(2760 series)*
5.1 Exports
5.2 Imports
5.3 Balances

6. National quarterly accounts *(686 series)*
6.1 Income statement of jobs and resources (fob-fob)
6.2 Final domestic consumption of households
6.3 Costs and margins
6.4 Income statement of jobs and resources (cif-fob)
6.5 Gross fixed capital formation
6.6 Imports and exports (cif –fob)
6.7 Imports and exports (fob –fob)

6.8 Labour units and income
6.9 Value added at basic prices
6.10 Value added at market prices

7. Labour force quarterly survey *(317 series)*
7.1 Labour Force
7.2 Non labour force
7.3 Employment
7.4 Population
7.5 Unemployment
7.6 Seasonally adjusted data
7.7 Rates

INDICATORS AVAILABLE IN THE CONISTAT DATABASE	**NUMBER OF TIME SERIES**				**CLASSIFICATION**			
	Raw	*Seasonally adjusted*	*Working day adjusted*	*Total*	**NACE rev.1.1 (ATECO 2002)**	**MIGS**	**Coicop 95 Rev. 1**	**Other classifications**
Consumer price index	298			298			X	
Industrial output price index on domestic market	163			163	X	X		
Retail Trade sales index	402	3		405	X			X
Labour input variables in large firms (over 500 employees)	3312	6	2	3320	X	X		
Contractual wages	251			251	X			X
Industrial production index	360	24		384	X	X		
Industrial turnover index	420	12		432	X	X		
Industrial new orders (flow)	255	3		258	X			
Industrial new orders (stock)	255			255	X			
Foreign trade statistics	2760			2760	X	X		
Tourism activity indicators (arrivals, nights, average stay)	27	18		45				X
Quarterly national accounts	312	374		686				X
Labour force quarterly survey	285	32		317				
TOTAL	9100	472	2	9574				

5. FUTURE DEVELOPMENTS

- Data Sender (already on testing): a system that automatically and periodically send by e-mail to the user who subscribed this service, data sets of time-series, in Text or XML format, ready to be loaded in analysis software like SAS or Excel;
- Set larger time-series (if the Service Unit has got the statistical data);
- Include new simple computing capabilities on the time-series (e.g. the calculation of percentage changes);
- Enlarging the domains and sub-domains with other indicators referred to: agriculture, constructions, harmonized consumer price index, hourly contractual wages, working days adjusted industrial production index, etc.

- - - - -