The ECB Statistical Data Warehouse: improving data accessibility for all users

Gérard Salou[1]

Introduction

The Directorate General Statistics of the European Central Bank (ECB) is responsible for the efficient delivery of statistics needed for the ECB's monetary policy and for the other functions of the Bank, the Eurosystem and the European System of Central Banks (ESCB). In addition, it is responsible for providing statistics to the interested public and to market participants.

In September 2006, the ECB released the initial version of its new online data delivery service via the internet, the ECB Statistical Data Warehouse (SDW). The SDW is designed to address a wide range of euro area statistics users, from one-time visitors searching for the most recent value of a specific statistic or single time series, to more frequent users such as market participants, journalists, analysts and researchers. It features robust and simple-to-use interfaces that provide access to numerous features and functions, designed to make data accessible to users regardless of their knowledge of IT systems or the intricacies of statistics.

The result has been to improve service to all European citizens and members of the world community interested in euro area statistics, while increasing the transparency of the data behind ECB monetary policy decisions. Some 20,000 public users per month access the new SDW internet portal. This compares with only slightly over 200 users weekly for the ECB's previous data tools.

This paper begins with a description of the SDW's content and an explanation of the data and metadata model. It goes on to discuss the main features of the internet portal, focusing on their objectives with relation to the different categories of users they are intended to serve. Finally, information on future development is provided.

Data content

The SDW contains all euro area economic, financial and monetary statistics published by the ECB and Eurostat relevant to monetary policy. In particular, the SDW provides all euro area statistics published in ECB statistical publications, national contributions to euro area statistics calculated by the ECB and published jointly as part of the Eurosystem joint dissemination framework, and access to euro area national statistics or other national statistics published in ECB publications.

In principle, ECB data are available on the SDW at their official, pre-announced release time. In practice, there may be a few minutes' delay for Eurostat data, due to the transfer time between Eurostat and the ECB.

---

[1] European Central Bank, Directorate General Statistics, Kaiserstrasse 29, D-60311 Frankfurt am Main, Germany, gerard.salou@ecb.europa.eu.

The full history of revisions is available only internally at the ECB's SDW, though it will be made available to the public in the course of 2007.[2]

Data and metadata model

Metadata consist of information regarding data that users and computers require to locate, retrieve, process, understand, present and analyse data. For statistical data as well as for metadata, the ECB uses the SDMX[3] standard data model. In IT terms, and in terms of data, the SDMX standard is similar to a "star schema". In that model, datasets are organised in hypercubes (or "data cubes"). The dimensions of the different data cubes can be shared across datasets to make it possible to relate data across datasets.

The SDW organises and stores metadata using the following model and terminology:

- *Metadata types* are classified according to an extension of the IMF Data Quality Assessment Framework (DQAF) metadata classification. Additional metadata types are required for some ECB internal processes (eg publication);

- *Metadata elements* are the values that metadata assume, eg a string "excluding shares";

- *Metadata item*: a combination of a type and an element;

- *Metadata*: a combination of a metadata item and any specific part of a data cube (data and metadata item combined).

This model prevents redundancy of "metadata elements" which need to be present in the database only once, labelled with a type and attached to data, when necessary. Also, the model, based on its construction, makes it possible to attach "metadata items" to any level and part of the data structure, eg to an entire dataset, to a particular dimension or to any combination of dimensions, down to the finest granularity, the data point. This data and metadata model provides secure storage of all types of statistical data and metadata that users of ECB statistics need for reporting and analysis.

Technology

The SDW is built by incorporating this data and metadata model in an Oracle database. The system allows the SDW to be fed automatically by statistics-generating systems. Procedures allow easy inclusion of new statistical datasets and new types of metadata when required. The system also provides for timely data delivery and standardised update histories.

Internet users can interact with the Oracle database through a browser. The interface is equipped for use by different browsers and operating systems.

---

[2] Real Time Data Base.

[3] See http://www.sdmx.org.

Types of users

The SDW is designed to address a wide range of users of euro area statistics, from one-time visitors searching for the most recent value of a specific statistic or single data-series, to more frequent users such as market participants, journalists, analysts and researchers.
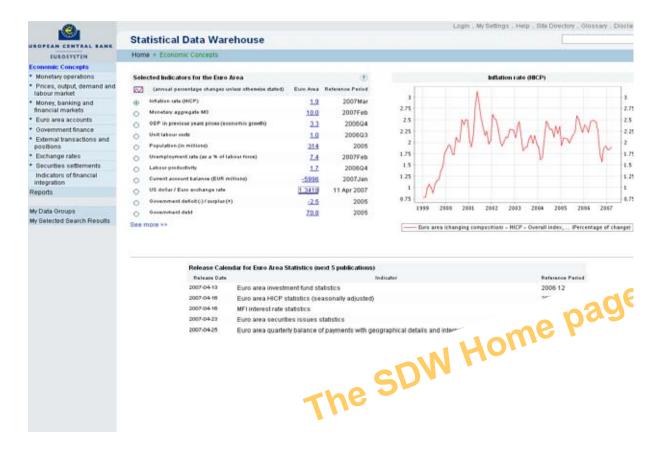
The user interface has been devised to facilitate access regardless of users' knowledge of euro area statistics, their level of patience with database navigation, and their knowledge of IT tools generally available on the internet for exploring databases.

The SDW has been developed in close collaboration with users, while the ECB internal version was developed with ongoing feedback from a representative user group. For the external version, two different usability tests have been conducted with users representative of relevant professions. These tests have made it possible to significantly simplify the user interface.

Features, by level of expertise

Users searching for euro area data on the internet and unaware of the SDW and ECB websites

For these users, we have designed the SDW so that all its html pages have a URL and, consequently, can be indexed by internet search engines. Searching for "euro area data" and including the economic category of interest generally returns the relevant pages of the SDW in the first page of search engine results, provided that the category being searched appears in the SDW. Users generally enter the SDW at the statistical publications page or at the Quick View, because rich text descriptions lead to those pages.

The SDW Home page

Users looking for statistics on the ECB web site who are not familiar with the SDW

These users will come across links to the SDW in many places in the Statistics section of the ECB website, in particular on the "key euro area indicators" page, which is constructed automatically from the SDW and includes links to all time series contained on the page.

Users looking for key euro area statistics who are unable or unwilling to navigate the database

The SDW homepage contains an interactive table and chart covering the most frequently requested euro area statistics. Clicking on the radio buttons updates the graph, and clicking on the figures displays a page with the corresponding complete time series, a graph, and the most important metadata associated with that time series. This page is referred to as the Quick View. Each time series contained in the SDW has a Quick View page with a unique URL. The Quick View metadata material includes references to the precise positions of the corresponding time series in the ECB publications in which they appear, along with a one-click link to the appropriate page of the relevant electronic publication. A number of basic transformations can be performed from the Quick View window (eg growth rates and frequency conversions).

Users able and willing to navigate a database

These users may want to navigate hierarchically by economic categories. They will be able to navigate to and select any data contained in the SDW. Time series selected are always accompanied by a reference to the publications in which they appear, so that users can see the context of each series and relate what they select to what has been published. After making the selection, users can view data in tables and charts, perform basic processes, view the metadata and export data to Excel, CSV, SDMX-ML, Excel Pivot tables or PDF formats.

Users familiar with ECB publications or looking for references to publications

ECB statistical publications (Euro Area Statistics Section of the Monthly Bulletin and the Statistics Pocket Book) can be accessed within the SDW. Users can see either updated PDF pages of each publication or the corresponding selections from the database, ready for processing or download.

Users familiar with IT systems but not with statistical classifications or publications

These users may want to use the SDW's search engine, which searches most metadata and returns time series. The algorithm uses a thesaurus, detects typographical errors and suggests alternatives to help users with statistical jargon. Search results are ranked by relevance. For the most part, the relevance of a time series is determined by the number of times it appears in publications.

Frequent and expert users of the SDW

These users may very well use all of the above, in addition to the data basket function, which they can use to customise their access to data and create and store their preferred data groups.

Usage statistics

The number of distinct users accessing the SDW since its launch in September 2006 has increased from 10,000 per month in September 2006 to approximately 20,000 in March 2007. This represents 100 times higher use than occurred with the previous version of the ECB online data service. The most frequently accessed pages are the Quick View pages accessible from the home page. Feedback from various categories of users has been quite positive. The significant increase in the number of users of euro area statistics has also led to a moderate increase in the number of questions related to data content. By contrast, users ask very few questions on the system's functionality.

Next steps

The next steps envisaged are to further enhance access for two important categories of users: those interested only in the main statistics, and frequent and expert users.

For occasional users, the Quick View and the home page will be improved to include greater functionality in the interactive graphs, so as to provide for more detailed analysis of the underlying data.

For frequent and expert users, the intention is to provide direct connectivity to the SDW from analytical applications using web service technologies.