Dynamic data presentation and interactive publication for official statistics

Hans-Joachim Mittag[1]
University of Hagen, Germany / Eurostat, Luxembourg
joachim.mittag@cec.eu.int

Keywords: Data dissemination, official statistics, interactive visualisation, Java applets

Summary

In today's global information society statistical offices have to cope with growing demands for user-friendly data dissemination and presentation. This paper describes an approach which aims at meeting these challenges by making use of interactive data visualisation. The visualisation is performed by means of self-contained and multilingual Java applets developed within the framework of a publicly funded German multimedia project "New Statistics". These elements, brought together offline or online in suitable repositories, could be shared between different statistical offices.

The paper also describes and exemplifies the idea of incorporating interactive components into electronic publications in official statistics. Such interactive publications enable the reader to spontaneously "play" with the data and to discover the main messages behind the figures by changing the data display or by visualising self-selected data subsets.

Introduction

Due to advances in computer and internet technologies, the dissemination policy of national as well as of supranational statistical offices has changed considerably during recent years. There has been a transition from a chiefly paper-based distribution of data sets and publications to an electronic distribution via the Internet. The few remaining key publications in printed form are complemented by CD-ROMs or DVDs. This development is driven by growing user expectations with regard to the immediate accessibility and comprehensibility of data as well as by policy targets like "Towards an Information Society for All".

In spite of the already existing technological capabilities, the potential of the new dissemination channels is barely exploited in official statistics. The Internet is mainly used for transporting static information with rather limited added value compared to printed material. Interactivity seldom means more than the provision of navigation tools supporting queries or extraction from databases. Flexible tools providing immediate data visualisation and interactive publications offering environments suitable for a quick exploratory ad-hoc data analysis are still largely missing. Such tools and user-friendly environments address citizens at large and offer the unique opportunity to increase the visibility of official statistics in society and to improve its public image. In addition, they could play an important role in furthering statistical literacy and in linking the rarely connected worlds of official and educational statistics.

Dynamic data presentation for press conferences, in-house seminars and on CD-ROMs

In order to illustrate the huge potential for official statistics connected with dynamic data presentation and interactive visualisation, several Java applets based on Eurostat data sets have recently been developed within the framework of a German multimedia project "New Statistics"

---

[1] The views expressed in this paper are those of the author and may not in any circumstances be considered as stating an official position of a statistical office or any institution of the European Commission.

([1], [2]). These elements are platform-independent, self-contained with built-in methodological comments, accessible free of charge and do not require to download additional plug-ins. They aim at communicating the messages behind the data by means of self-explanatory, user-controlled graphs to a broader audience. Such components could be made available in the form of multilingual repositories and might be shared and jointly maintained by different national or supranational statistical offices.

One possible application of the elements of such repositories is their integration into PowerPoint slide shows employed in press conferences or training courses. Figure 1 shows a Java applet displaying German results from the recent Structure of Earnings Survey (German SES 2001). The applet was designed for a press conference organized in autumn 2003 by the German Statistical Office for the presentation of key survey results to the media. One may change the data subset to be graphically displayed via predefined menus, slot in the numerical values behind the graphs or open a second window giving access to methodological comments, for example on the European classification scheme NACE for economic activities. Part (a) of Figure 1 displays bar charts showing the monthly earnings of all employees for a user-selected NACE section or a NACE aggregate without regional breakdown. The user has the option to switch to a subset of employees, to annual earnings or to the former West or East Germany. Part b refers to only one NACE section. The user may now alter in a second step the economic activity or, again, the reference population or the reference period.

*Figure 1: Dynamic presentation of results from the German SES 2001*
a) Data for different NACE sections   *b) Data for a user-selected NACE section*
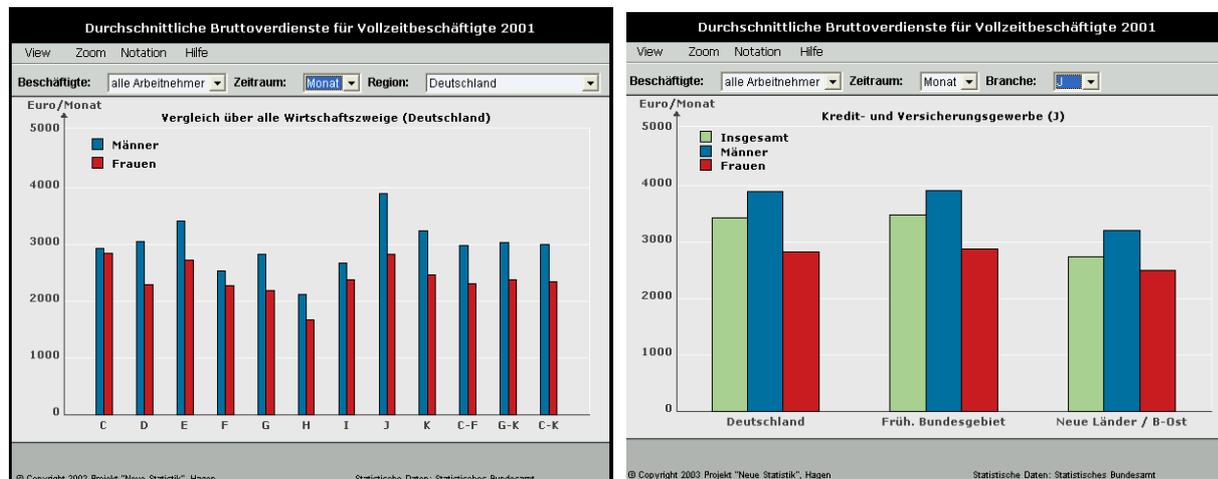
Java applets could also be supplied on CD-ROMs or DVDs as a complement to traditional print publications. Figure 2 shows a trilingual interactive element which has been incorporated into the CD-ROM complementing Eurostat's Regional Yearbook 2003 ([6]). The user can select a set of capital-city regions and compare the values related to a socioeconomic variable chosen from a menu. Alternatively, the user may simultaneously compare two variables but in that case for only two regions. Detailed methodological comments including a map could be faded in and the default language English is alterable to French or German.

Figure 2: Interactive presentation of socioeconomic variables for European capital-city regions
(the map is part of the methodological comments to be activated by clicking on "Notes")
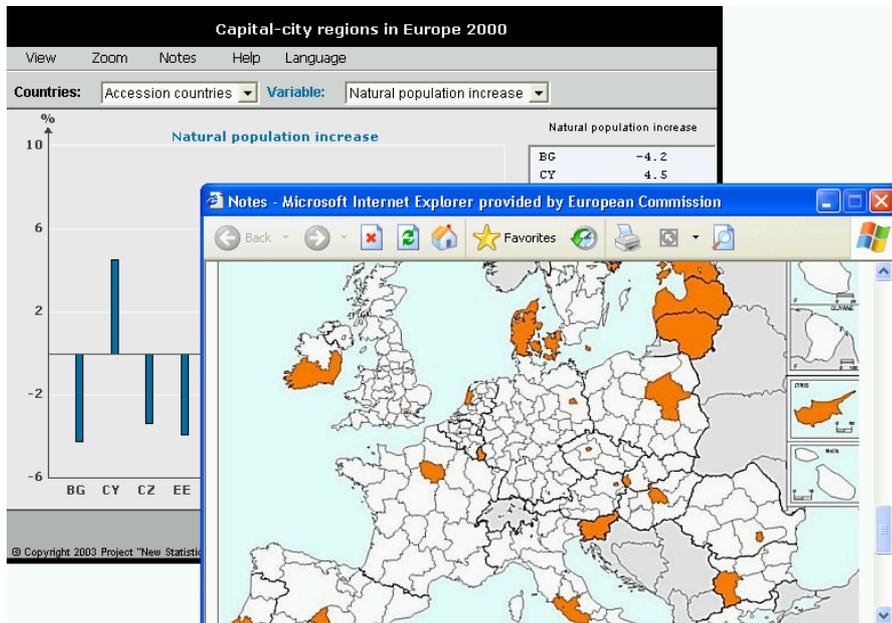

Virtual libraries and interactive publication

Repositories with self-contained interactive components for dynamic data presentation could also be implemented online. The Java applets presented in Figures 1 – 2 and further interactive elements are, for example, already accessible via Internet. Most of the elements presented there (see [3]) offer the option to choose between different graphical instruments, such as bar charts, time series graphs or boxplots. Such virtual libraries, possibly containing contributions from different content providers and enriched with suitable meta information, might contribute to tying together the worlds of official and educational statistics.

User-controlled visualisation tools might likewise be used for creating interactive publications. These connect static text with user-friendly working environments for exploratory ad-hoc data analysis. Part a of Figure 3 shows an excerpt from an experimental interactive publication (accessible via [4]) dealing with results from the last European labour costs survey (LCS 2000). Within the hypertext, the reader may activate links to Java applets and also to a glossary or to traditional pdf-type web documents. After starting a Java applet (see link at the bottom of Figure 3a), the applet appears, similar to the map in Figure 2, in a new window. As a starting-point, the applet default provides exactly the same graph as displayed in Figure 3a but one may now alter the default and select another pair of countries from the menu. This is shown in part (b) of Figure 3. Another option is the comparison of labour costs for two economic activities over a set of countries. In addition, the user may wish to fade in the numerical values behind a graph, as illustrated in Figure 3c. Finally, the applet offers alternative graphical tools for exploratory data analysis, such as boxplots. The extreme values of the data set defining a boxplot are identifiable by putting the mouse on the upper or lower part of the graph. Figure 3d shows this for a boxplot referring to data from Cyprus where "Construction" (NACE section F) is associated with the highest hourly labour costs.

*Figure 3: Interactive publication dealing with results from the Labour Costs Survey 2000*

a) Excerpt from the interactive publication with hourly labour costs for two countries
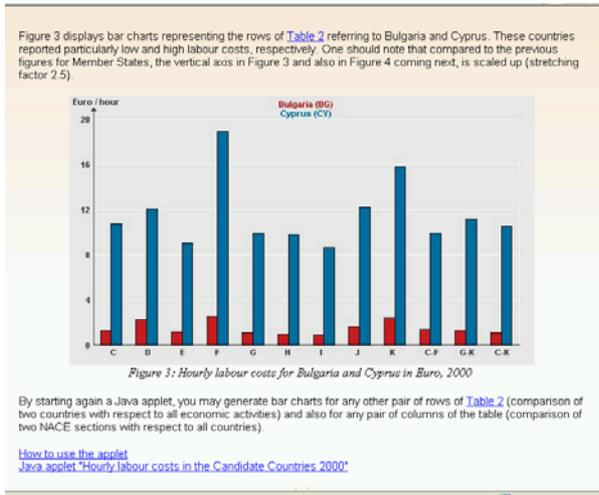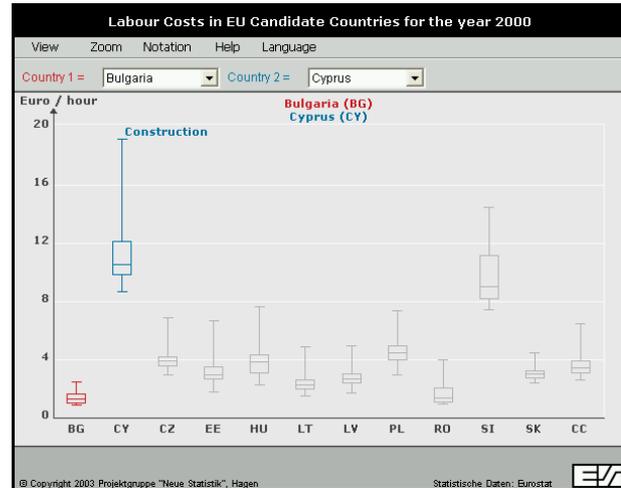
b) Window opening on activating the applet after changing the default selection of two countries
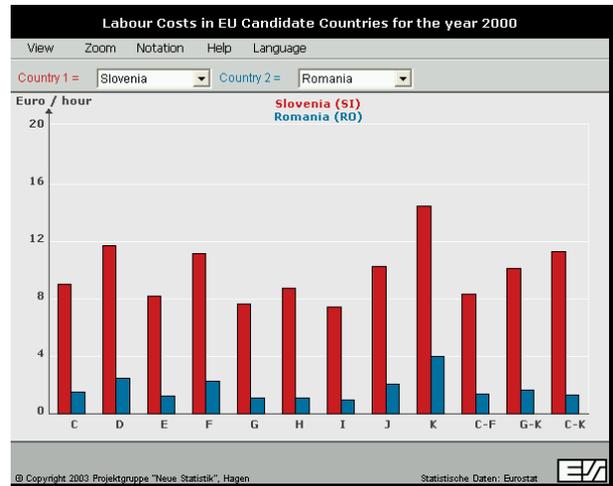
c) Java applet after changing the data display (comparison of two NACE sections over countries)
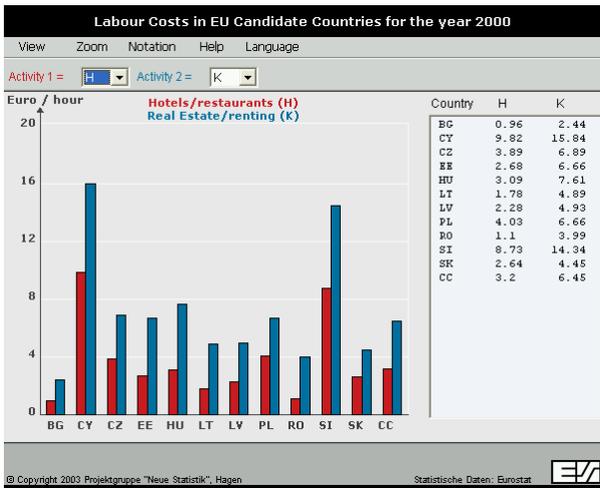
d) Applet after changing the graphical instrument (boxplots instead of bar charts)

Outlook

In the near future, interactive visualisation tools will increasingly often be directly connected with large databases. The Java applets presented above are still not linked with a database but the development of suitable interfaces is nowadays feasible without any substantial technological problems. Such a linkage would ensure an automatic update of bar charts or other graphical instruments. Eurostat's new dissemination policy is a move in this direction and offers, free of charge, a growing number of database-supported tables with standardized meta-data and with optional visualisation.

Another rapidly developing area is the presentation of regional data by means of interactive maps. Suitable technical platforms are either based on Java (see [5], for example) or on SVG (scalable vector graphics), a new XML-based standard for vector graphics on the web. A few NSIs, for

instance those of Germany, the United Kingdom and Canada, already use interactive maps on an experimental basis (see [7]).

The potential associated with interactive publication is still unexploited in official statistics but the added value of the concept compared to traditional publication modes is obvious. It is already discernible that commercial solutions for web-based online reporting and dynamic data presentation are gaining ground. However, such solutions require licensing for supplier-specific browsers.

Web links and other references :

[1] http://www.neuestatistik.de
[2] http://www.neuestatistik.de/english.php?HP=4
[3] http://forum.europa.eu.int/Public/irc/dsis/wages/information (click on "Java applets ..")
[4] http://www.fernuni-hagen.de/e-statistik/publikationen (click on "Labour costs ..")
[5] http://www.statistik-bw.de/intermaptiv/home.asp
[6] Regions: Statistical Yearbook 2003, Office for Official Publications of the European Communities, Luxembourg, 2003.
[7] Smith, A. / Neutze, M. (2003), "Intelligent Graphics"– SVG for website data visualisation, Paper presented at the 54th Session of the International Statistical Institute, Berlin.