Seminar on Dynamic Graphics for Presenting Statistical Indicators

5-6 March 2007
ISTAT headquarters, Rome, Italy

SUMMARIES OF PRESENTATIONS
Interactive Visualization of Statistical information

Hans-Joachim Mittag, University of Hagen, Germany

Statistical information has a pertinent importance for evidence-based policy-making at national and international levels. Data, in particular data aggregated to indicators, are used to develop and to support strategies and action plans as well as to monitor and assess their performance. The huge demand for policy-relevant and reliable statistics is accompanied by a growing demand for easily accessible environments for the presentation and dissemination of data and policy performance measures.

Today’s advanced state of information technology has opened up opportunities to meet these challenges and to communicate statistical information in a way that can be easily understood not only by experts. User-friendly, interactive visualisation tools are meanwhile available for presenting and communicating official statistics. This statement will be demonstrated by means of an interactive Java-based visualisation tool developed in connection with a German multimedia project for statistics education. The tool is platform independent and self-contained with built-in methodological comments and a help functionality. It aims at communicating the messages behind statistical information at first glance by evaluating user-controlled graphs. In the European Commission, the tool has been successfully applied for presenting, amongst others, the structural indicator “European unemployment rates”, the main results of the recent European structure of earnings survey, and the “European e-business readiness index” which is related to the e-Europe action plan. It has been applied on a CD-ROM accompanying Eurostat’s Yearbook on Regions for disseminating statistical information on characteristics of European capital-city regions. Furthermore, Java-based interactive elements have been incorporated into selected Eurostat online publications in order to illustrate the added value of this approach.

The interactive elements can be used offline on standard data carriers (CD-ROM, DVD) or as part of a PowerPoint presentation, as well as online embedded into virtual libraries or as an e-mail attachment. The design aims at enabling the educated lay public to explore a data set from different perspectives by using different graphical tools (bar charts, time series graphs, scatterplots, boxplots) and by evaluating different sub-sets of the data universe (breakdown by regions, by economic activities or by sex). The mouse can be employed for slotting in additional information (numerical values, codes for country or economic activities).

An important design principle was to strictly avoid information overload or unnecessary visual distraction. This implies in particular to ensure lucidity by restricting the number of simultaneously presented graphs, to cautiously employ colours and to abstain from superficial show effects.

Links:

http://www.fernuni-hagen.de/newstatistics
Provides information related to a German multimedia project “New Statistics”. The path “Library .. / Official statistics” gives access to a virtual library containing Java applets visualizing official European data sets. Recent conference contributions dealing with interactive data visualization are accessible via “Project related publications”.

http://forum.europa.eu.int/Public/irc/dsis/wages/information
The second link gives access to a web site of the European Commission containing Java applets based on Eurostat data. The first link leads to interactive online publications published within Eurostat’s series “Statistics in Focus” (experimental editions).

http://statind.jrc.it/ebusiness.html
The second link of this web site of the European Commission gives access to the Java applet “European e-business readiness index” mentioned above.
Java 2D High-chart

Christian Renwanz, Deutsche Bundesbank

In March 2000 two developers from Bundesbank and three from the company GFT started to develop the new interactive charting tool “HighChart”. It shall replace the mainframe system called “TWPLLOT”, used till today for most publication charts of the Deutsche Bundesbank.

Special features of Bundesbank Charts
- Overlapping logarithmic scale-areas on the y-axis
- Sub charts with common logarithmic scale factor to keep comparability between them
- Flexible axis data model: equidistant month- or day scale with arbitrary calendar-models (“All Days”, “Monday to Friday”, “Working Days”...)
- Need for Automation (e.g. monthly reports)
- High Quality (e.g. Adobe type 1 hardware fonts)

A sample chart
This chart has three sub charts. The sub chart in the middle has a linear scale, the other two a logarithmic scale with a common logarithmic scale factor (notice the distance on the paper of the labels 1,30 to 1,90 in sub chart one is the same as the one in sub chart three between 130 to 190!)
The scale areas of sub chart two and three overlap; the horizontal gridlines have to be cut.
The x-axis data model is “equidistant months” nevertheless daily values are plotted for the last year of the x-axis.

Highlights of HighChart
- Interactive tool to generate time series charts in high quality for publication purposes
- Date axis models with equidistant days or months and arbitrary ticks and labels
- Logarithmic axis models with the possibility to lock the scale factor with other axis
- Periodicity of data (e.g. monthly data) to be represented and x-axis data model (f.e. “Working Days”) free combinable
- User coordinates (x=01.03.1996 and y=150 for the label “USA” in the sample chart)
- Blanking (used to make labels on gridlines more readable, see also the label “USA” in the sample chart)
- Template mechanism (Dynamic linkage of charts or properties)
- Scalable Vector Graphics (SVG) and PostScript export, own PostScript driver (using Adobe type 1 hardware fonts)
- XML file format and the possibility to create own drag & drop resources in XML
- “MiniWord” for text processing with multicolumn text and rendering of Adobe type 1 fonts on the screen.

Technologies and Tools
- Eclipse, Borland JBuilder and Together J
- Java 2D API and JDK 1.5 from SUN
- JTEST, JPROBE
- ANT (xml based build tool)
- Data access to Bundesbank database: RPC-call via JDBC:ODBC bridge
Adding Animation and Interactivity to National Statistics

Alan Smith, Office for National Statistics, UK

ONS are pioneers in the use of SVG (Scalable Vector Graphics) for interactive data graphics. Our interactive population pyramid template (for time series display of population statistics) has been established for over 3 years and has been openly shared for use by other NSIs across the globe (e.g. Australia, Norway, Czech Republic, the Netherlands, Ukraine).

A key element of ONS’ work in this area has been to focus on design principles for interactive maps and graphics and how best to incorporate the elements of animation and interactivity. Exploring the use of interactive graphics to support commentary on patterns in datasets is another theme to our work.

ONS’ outputs including separate, stand-alone graphics to support individual datasets (e.g Urban Audit, Area Classifications) as well as on-the-fly generation of XML data on large scale web services (Neighbourhood Statistics). A wide variety of indicators have been included in this work, both social (inc. deprivation) and economic.

This presentation would cover ONS’ experience in this area, user feedback, technical implementation and a discussion of the design principles of interactive graphics.

References – example graphics:

Interactive Population Pyramid time series in SVG
www.statistics.gov.uk/populationestimates/svg_pyramid/default.htm

Area Classification mapping in SVG
http://www.statistics.gov.uk/about/methodology_by_theme/area_classification/la/svg/index.html

Urban Audit II mapping in SVG
http://www.statistics.gov.uk/urban_auditII/urban_auditII_map.htm

Neighbourhood Statistics service, featuring dynamic SVG
www.statistics.gov.uk/neighbourhood
The Statistics Netherlands Business Cycle Tracer; Visualising the State of the Economy

Floris van Ruth, CBS, Netherlands

The SN Business Cycle Tracer is a web based application which combines sophisticated economic analysis with an advanced presentation. Our aim was to develop a tool which could unify and clarify the information on economic developments published monthly by Statistics Netherlands. In addition to the normal publication of short-term economic statistics, we wanted to make the state of the economy visible, in an easy to interpret manner. This was achieved by creating a graphical representation of the cyclical development of a selected set of important macro-economic indicators. By focussing on the business cycle component of the statistics, a lot of noise was eliminated and the dynamics of the economy became more clearly visible. The graphical representation is a diagram divided into four quadrants, corresponding to four phases of the business cycle. The placement of the economic indicators in the diagram then results in an easy to interpret representation of their development and of the state of the economy as a whole. The indicators move around the diagram as the business cycle develops. This visualisation makes a sophisticated economic concept easy to understand. The application is animated, which means that it is possible to play back the past developments in the economy. It is also possible to select one or a few indicators to study their past development and interrelations in greater detail. This is an innovative way to report on the state of the economy, with dynamic features which are rare in the publication of economic statistics.

Link:

Business Cycle Tracer
Rich Internet Applications (RIAs) are a hybrid between desktop and web-based applications and aim at offering the best of both worlds: the rich user experience and responsiveness of desktop applications together with the low maintenance costs, platform independence and broad reach of internet applications. Several technologies are available for implementing RIAs, such as Ajax, Flash (Flex, OpenLaszlo) and Java.

The Flash Player offers many advantages for developing RIAs. Independent research companies regularly show that the Flash Player is almost ubiquitous and that new versions are quickly adopted. It offers a rich set of components for building user interfaces as well as a comprehensive charting library. The data can be fed to the Flash Application using plain XML data files (such as the SDMX-ML data files published by the European Central Bank), web services and Java objects. Last but not least, most of the underlying technologies (ActionScript, which is ECMAScript compliant, CSS, DOM Level 3 Events, E4X, XML, etc) and design principles (e.g.: the MVC pattern) are already familiar to most web developers.

The European Central Bank plans to use the Adobe Flex technologies to improve the visualisation of statistics published on the ECB website and the Statistical Data Warehouse. The two prototypes that we would like to demonstrate cover two different use cases:

1. The first demo will show a new version of the graphs for the bilateral exchange rates and aims at demonstrating how RIAs can help improving the visualisation of statistical data. Similar improvements as the ones that will be demonstrated are also planned for the Statistical Data Warehouse.

2. The second demo aims at demonstrating how RIAs can help to improve the understanding of concepts behind statistical data, using a scenario-based or "story-telling" approach. It will show the evolution of the Harmonised Index of Consumer Prices (HICP) in the euro area. The prototype offers ways to visualize data for all countries of the euro area and all major components of the HICP. It will also allow a user to set his consumption habits and see the inflation index derived from them.
Dynamic Data Displays: Experiences from USDA's Economic Research Service

Vince Breneman, USDA, Economic Research Service

The U.S. Department of Agriculture's Economic Research Service (ERS) informs and enhances public and private decision-making on economic and policy issues related to agriculture, food, natural resources, and rural development. To accomplish this task, ERS economists and social scientists develop and disseminate economic and statistical data to the public through the web. Though web-based technologies provide great opportunities in their ability to disseminate research and policy information, with access to over 9000 datasets, a broadly defined research program and a diverse clientele, the development of standard, flexible and maintainable data delivery products creates significant challenges.

From time series to geospatial to mixed spatiotemporal data displays, we will describe our various approaches, tools used, functionality considerations, constraints and tradeoffs. This presentation will highlight experiences developing dynamic data displays from three of our more popular applications: The Food Consumption data system which provides time series displays of the amount of various types of food available for U.S. consumption; The Profiles of America web-based mapping application which allows one to map key socioeconomic measures and summarize data across a series of regional typologies; and, The Agricultural Resources Management Survey application, our most complex application covering displays of mixed spatial and temporal data quantifying the financial condition, production practices, resource use, and the economic well-being of America's farm households.

Link:

Data Sets – Featured Products
http://www.ers.usda.gov/Data/
Agricultural Trade by Origin and Destination

Edward Gillin, FAO

The FAO Statistics Division will make a presentation showing dynamically agricultural trade by origin and destination (WATF). This highly informative animated demonstration of world and national trade, vividly portrays the inter-relationships between and among countries. It can identify the primary origin of the products which are often repackaged and re-exported to third countries. It is fully supported by another application (WATM) which is a numeric trade data matrix. This presentation can also be applied to trade in water (WATW) via examining the water requirements of products and is particularly interesting for regional water resource analysis. Therefore, the presentation to be made will cover the WATF, WATM and WATW.

Link:

Core Trade Matrix – (WATM)
Presenting OECD Main Economic Indicators as SVG-maps using PX-Map2

Marianne Vik Dysterud, Statistics Norway

PX-Map2 provides an overview of your data in a simple and efficient way. The program visualizes your statistics in dynamic graphics like choropleth maps, symbol maps, data distribution chart, pies and bar chart. Adjustments of colors and classification type are made on the fly and stored in cookies. Scrolling values of a variable could be experienced as an "animation". Other functionalities are zooming in/out and panning. PX-Map2 also has possibilities for 2-colorscales around a user-defined threshold-value.

Adoption of your own needs for language, default-settings for colors, classifications-type etc. is easy done in a text-editor.

PX-Map2 runs as a standalone program or could be launched from another program (ex. PC-Axis). It runs on local PC and web server.

Data requirements: A basic map in SVG format, statistical data stored in PC-Axis file format or character-separated text file (i.e. semicolon separated).

Application of GIS technologies to perform dynamic thematic mapping on the Web

Pina Grazia Ticca and Stefano De Francisci, ISTAT, Italy

Statistical indicators can be mapped and dynamically represented on digital cartography on the Web, mainly following two different strategies. The first one is oriented to migrate on the Web traditional Client/Server systems that performs dynamic thematic mapping, keeping in that way their strength and complexity. The second one is what we can call “without-server solution”, strongly oriented to create and disseminate compact, immediate and easy-to-use dynamic thematic maps and graphs, based on small static datasets, but not fully exhaustive.

Each of these strategies has its own advantages and disadvantages, both in designing and development, and in usage and interaction.

GIS technologies nowadays are powerful enough to support both solutions. This presentation intends to show, through interactive examples, what we have done and what we’re prototyping at the moment with the two strategies, in order to efficiently answer to the variety and different profile of Internet users.

References to the first on-line system:

14th General Population and Housing Census – Legal Population

http://cartema.istat.it/pop_leg_eng/center.asp?Copertura=0&Livello=0&tema=tema10pl&hval=1024&wval=1280&q=pl01000100112001

Municipalities by their maximum peopling period (1861-2001)

http://cartema.istat.it/pop_leg_eng/center.asp?Copertura=0&Livello=0&tema=tema10pl&hval=1024&wval=1280&q=pl01000100112001

Resident population density (inhabitants/kmq)

http://cartema.istat.it/pop_leg_eng/center.asp?Copertura=58&Livello=3&tema=tema3pl&hval=1024&wval=1280&q=pl02000105842001
Most issues that governments engage with - such as health, education, economic performance and crime - involve multiple variables that interact in non-obvious ways. For informed citizenship, and personal well being, citizens (and politicians and journalists) need to be able to reason with complex multidimensional data. Everyone concerned with the display of data has a responsibility for the education of users. There is a large literature on the problems that students and adults have with simple concepts, such as interpreting static 2D graphs, and tabular information. Statistics packages require a high level of statistical knowledge before they become useful (and are often ill-suited to the analysis of non-linear relationships, or where variables show their effects over a limited range).

However, our recent studies show that students aged 9 – 14 years can reason with multivariate data, if the data are presented on screen, and they are given control of what is displayed. Our empirical work (e.g. Ridgway et al., 2006) shows that computer-based multivariate tasks are no more difficult than 2D paper based tasks. We are developing a number of software ‘shells’ in Flash that run on any web browser, available as ‘freeware’ downloadable from our website. They allow up to 5 variables to be displayed under user control. These displays are being integrated into curriculum materials for students aged 12-15 years.

The quality of an interface can best be judged by analysing the interpretations that representative users make of data sets – we believe that all other evaluative methods (analyses of transactions and other ergonomic analyses), though important during design, are surrogate measures.

We will show some interfaces, and would welcome discussions on possible collaborative ventures with seminar participants.

Link:

Smart Centre
http://www.dur.ac.uk/smart.centre/
Eurostat approach to graphical representation of statistical data

Gunter Schäfer, Thomas Werkhoven - Eurostat

The value of statistical data for users can be significantly enhanced through adequate graphical tools for locating relevant statistical information in vast statistical databases as well as displaying statistics. The paper describes the principles Eurostat applies in designing such tools and introduces the key projects of Eurostat which are currently in the implementation phase. A high degree of interactivity of the user with the Internet in order to tailor the information according to specific needs as well as the integration of different types of information in particular statistical data and methodological data are two key elements of the approach. The key projects introduced cover a tool for display and manipulation of tables, graphs and maps, a country and regional profile tool, a new highly interactive user interface to databases, as well as a tool for analysing the business cycle.

In the Eurostat approach for the access to statistics, a navigation tree that resembles the Windows explorer plays an important role. The user is invited to successively open the hierarchical structure until he/she reaches the data of interest. The data is organised in data sets that contain statistics which are closely related, i.e. one or more indicators that cover a certain statistical topic. In total the Eurostat databases contain about 3 to 4000 different data sets. Eurostat distinguishes between open datasets which are multi-dimensional in nature and predefined tables that present the most important statistics in a two-dimensional form for easy access.

In a first stage, the most efforts to present statistics in a graphical or map based form concentrates on the predefined tables. The Tables-Graphs-Maps tool treats the table view on equal footing to the maps and a graphical view. Maps and graphs are particularly suitable to compare data, e.g. to present the data in descending order, compare different countries etc. The user is provided with a range of functionalities for this purpose.

Closely connected is the Country Profile and Regional Profile tool. The user displays a wide variety of indicators per country and compares countries and regions with each other or with the European Union. This tool overcomes the general indicator orientation of the Eurostat statistical databases by a country or region orientation. Again a series of functions assists the user in graphical and map based analysis.

A technically more demanding project is the New User Interface project that aims at replacing the different interfaces for the extraction of data used in Eurostat by a single tool. The idea behind this interface is that once the user opens a data set, he/she sees a default view that is interactively modified in order to gradually arrive at the desired view on data. Evidently, this view can then be saved for further usage. The graphical and maps based views will be gradually integrated into this interface.

The business cycle clock originates from an idea of the CBS Netherlands for displaying about 20 business cycle relevant indicators in a way that allows locating the values of the indicators within the economic business cycle. In cooperation with the CBS, Eurostat will enhance the scope of the tool to all EU countries and EU/Euro area aggregates.
Navigation tree access to simple predefined tables

Navigation tree access to interactive Table View for complex data sets

Interactive customizable Maps

Country and longitudinal comparisons

Interactive customizable Graphs

Country and Regional Profiles
Millennium Development Goals Dashboard

Jochen Jesinghaus, JRC, European Commission

The Millennium Development Goals, agreed at the 2000 United Nations Millennium Summit by 147 Heads of State and Government and 189 UN Member States in total, are the globally recognised political framework for development cooperation. The eight MDGs are: Eradicate extreme poverty and hunger, Universal primary education, Gender equality and empowerment of women, Child mortality, Improve maternal health, Combat HIV/AIDS, malaria and other diseases, Ensure environmental sustainability, Global partnership for development. They have been operationalised through a set of approx. 60 MDG indicators, defined in the MDG Roadmap (United Nations General Assembly, document A/56/326 of 6 September 2001) elaborated in the wake of the UN Summit by members of the United Nations Secretariat and representatives of IMF, OECD and the World Bank.

JRC's "MDG Dashboard" is a software tool that presents the Millennium Development Goals indicators in a highly communicative format (e.g. colour-coded World maps) aimed at decision-makers and citizens interested in the Millennium Development Goals and Sustainable Development. The tool was elaborated in cooperation with the Canadian International Institute for Sustainable Development (IISD), and was originally designed for the indicator set of the United Nations Commission on Sustainable Development. JRC provides the software for ca. 200 countries and 15 years (1990-2004), based on regularly updated original United Nations statistics. The tool enjoys a high popularity (with several hundred Google hits), since it allows a transparent access to over 100,000 data items that otherwise would be presented in the form of huge spreadsheets and tables, inaccessible to non-statisticians. In addition to the MDGs, the dashboard of sustainability software has been applied to numerous other indicator sets, including the European Commission’s Sustainable Development and Structural Indicators, the UN CSD set, governance indicators and several sub-national indicator sets, e.g. in Australia, Canada, Estonia, Greece, Italy, Mexico and the U.S.

Links:

MDG Dashboard
http://esl.jrc.it/dc/

MDG Roadmap

United Nations Statistics
Dynamic Maps for Presenting Statistical Indicators

Astrid Marschatz, DevInfo Project Manager, UN Development Group Office

Overview
DevInfo is a database system that harnesses the power of advanced information technology to compile and disseminate data on human development. In particular, the system has been endorsed by the UN Development Group to assist countries in monitoring achievement of the Millennium Development Goals (MDGs). DevInfo provides methods to organize, store and display data in a uniform way to facilitate data sharing at the country level across government departments, UN agencies and development partners. DevInfo has simple and user-friendly features that produce tables, graphs and maps for inclusion in reports, presentations and advocacy materials. The software supports both standard indicators (the 48 MDG indicators) and user-defined indicators. DevInfo is compliant with international statistical standards to support open access and widespread data exchange. DevInfo is distributed royalty-free to all Member States and UN agencies for deployment on both desktops and the web. The user interface of the system and the contents of the databases supported by the system include country-specific branding and packaging options which have been designed for broad ownership by national authorities.

Highlights
The vision that DevInfo supports is a day when member states use common database standards for tracking national human development indicators, containing high-quality data with adequate coverage and depth to sustain good governance around the agenda of achieving the MDGs. DevInfo is enabling the UN system in realizing this vision as a general purpose database system designed for the compilation, dissemination, presentation and advocacy of human development indicators. By serving as a common database, DevInfo is being used to add value to national statistics systems by complementing existing databases and bridging data dissemination gaps. DevInfo is also being used as an advocacy platform to engage all stakeholders in policy choices for human development.

Innovations Achieved
DevInfo 5.0 has evolved from a decade of innovations in database systems that support informed decision making and that promote the use of data to advocate for human development. A major innovation of DevInfo 5.0 is the introduction of data and metadata standards to encourage open access and use of data across multiple organizations, platforms, and systems. DevInfo has adopted international standards in the areas of indicators (SDMX ISO/TS 17369:2005), data sources (DDI/Dublin Core) and digital maps (ISO 19115:2003).

Lessons Learned for Replication and Scaling Up
The DevInfo initiative is being implemented under the endorsement of the UNDG in collaboration with more than 20 UN agencies. More than 10,000 professionals have been trained in the use of DevInfo for improved statistical literacy and database administration (approximately 60% government and 40% UN professionals). More than 80 national statistics organizations and other agencies have officially launched an adapted DevInfo database with their user-specified requirements. There are a number of UN Agencies that have published adaptations of DevInfo - namely ILO, UNHCR, UN Habitat, UNHCR and UNFPA. For the second consecutive year, the UN Statistics Division has published the official UN data on MDGs indicators in an adaptation of DevInfo, called MDG Info. MDGInfo 2006 has been prepared to accompany the Millennium Development Goals Report 2006, presenting the most up-to-date country-level statistics available in the UN as of July 2006 for the global monitoring of progress achieved towards the MDGs since 1990.
Dynamic Maps for Presenting Statistical Indicators
The DevInfo mapping module has achieved a technological breakthrough in the presentation of statistical indicators in dynamic maps. With a click of a button, time series data are assigned to maps over time and then animated to show dynamically the change of any given statistical indicator across time periods. This produces an attractive way to show change in statistical indicators over time. This is possible because the database system contains digital maps linked to each data observation stored in the database. These digital maps are, in turn, linked to map metadata which indicate the representivity (start/end dates) of each digital map at national and sub-national levels. Then when maps are created on the fly, DevInfo automatically links the statistical observations to the correct maps by time period. For example, old census data are displayed in old maps with old administrative boundaries and new census data are displayed in new maps with all of the sub-national administrative boundaries updated over time.

Contacts
DevInfo  www.devinfo.org;  UNICEF is chair of the DevInfo Interagency Advisory Group;  DevInfo Global Administrator, Nicolas PRON  (npron@unicef.org);  UNDGO DevInfo Project Manager, Astrid MARSCHATZ  (astrid.marschatz@undp.org);  DevInfo Support Group  (support@devinfo.info).
SDMX and Open Source for cross platform graphics

Russell Penlington, OECD

Graphical representation of statistical data can be rendered and displayed independently of operating systems by using open source developing environments and international data exchange format standards. As an application of this, free Java charting libraries can be employed to display SDMX-ML data graphically. Java is an open source developing environment that can run on all operating systems which have the Java run-time environment such as Windows and Linux based systems like UNIX and Macintosh. Java based graphics can also be rendered on the web using Java server applets. SDMX is an ISO standard XML specification for exchanging statistical data and metadata between organisations and applications.