

**UNITED NATIONS STATISTICAL COMMISSION
and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

UNECE Work Session on Statistical Dissemination and Communication
(Geneva, 13-15 May 2008)

Topic: Communicating statistics in the information age

REFLECTION ON GRAPHICAL INTERFACES FOR ILLUSTRATING EUROPEAN STATISTICS

Submitted by Eurostat¹

Summary: This paper focuses on some of the challenges in displaying European statistics on the Internet. After a quick overview of data consultations on web site usage, this paper addresses a few aspects on the role of visualization tools for the web site. A recent survey shows that European National Statistical Institutes place significant value in illustrating statistics through interactive map applications. Together with the Member States, Eurostat has created a Task Force for exploring potential current and future graphical tools and increased co-operation. This paper outlines the first recommendations of the Task Force and further suggestions on making data visualisation more appealing.

I. INTRODUCTION

1. In October 2004 the European Commission established the new policy of providing all data free-of-charge through the Internet. Due to the new policy, the number of Eurostat website visitors per month increased up to 1 million. As outlined in section II, data consultations have become very popular; they are the cornerstone of the statistical website.
2. Besides the increased availability of the data, the way of presenting them has also undergone a change. Until recently Eurostat's data was only shown in a traditional tabular format. In November 2007 Eurostat launched the Table-Graph-Map (TGM) tool which, besides tables, also offers graphs and maps, and a toolbox for customization. First web traffic numbers indicate a significant demand for graphical displays: after the initial default table view of the TGM, up to 80% of the users enter graphs and/or maps. Eurostat attributes great value to graphical illustration of data as a supporting function of the core business of providing high quality access to statistics. A reflection on the role of visualization tools in the website can be found in section III.
3. A recent Eurostat survey on map visualization reveals that European National Statistical Institutes consider graphical means of displaying data on the web very important, albeit that the development costs are relatively high with a large variety of technical solutions and so far with relatively modest numbers of users. A Task Force on Visualization initiated by the ESS Dissemination Working Group has been launched in order to explore further steps and common approaches for developing and implementing graphical interfaces for statistics. Section IV highlights the first recommendations and further suggestions.
4. Conclusions are drawn – to some extent preliminary - in section V.

¹ Prepared by Thomas Werkhoven (national expert at Eurostat's Dissemination unit). The views expressed in this paper are those of the author and do not necessarily reflect the official policies of Eurostat. Special thanks for reviewing and improving the paper go to Gunter Schäfer (Head of Dissemination) and Francis Di Meo (User Support). E-mail: Thomas.Werkhoven@ec.europa.eu

II. DATA CONSULTATIONS AS THE CORNERSTONE OF THE WEBSITE

5. Before the introduction of the free-of-charge policy the Eurostat website attracted around 150 thousand visits² per month. After the introduction of the new policy in October 2004, the number of visits grew up to 1 million.

6. To a large extent the increase of the website usage is driven by the number of data downloads (up to 700 thousand). The most requested are the small and ready-made predefined tables that Eurostat provides for its key statistics. The user-defined data extractions are also highly demanded: the number of unique visitors extracting data from the data base averages around 50 thousand per month, roughly representing 1/5th of the total number of visitors. Although unregistered users account for the majority of downloads, the small number of external registered users (1%) renders very substantial flows of data in terms of numbers and size, and special services are in place to accommodate these needs.

7. The most popular fields of interest in terms of consulted data sets, web pages, search key words, and publications are concentrated on the key topics of economy and finance, population and labour and regional statistics. Besides these more aggregated types of statistics there is significant demand for detailed external trade data, especially by professionals. From log files analysis it furthermore appears that many Eurostat visitors are searching for country specific data, more or less similar to the demand for statistics on regions/neighbourhoods at the national level.

8. The user feedback and benchmark studies carried out show - besides general satisfaction - the need for further improvement of the navigation and access to data. The variety of data extraction tools and the number of steps needed for each extraction have been identified as critical points. On the other hand Eurostat's Navigation Tree – that is the hierarchical grouping of all data sets in a tree folder structure - is regarded as a strong point for finding data sets.

III. ON THE ROLE OF VISUALIZATION TOOLS IN THE WEBSITE

9. In November 2007 Eurostat introduced a new graphical tool for displaying predefined tables on the web; the Table-Graph-Map interface (TGM). Besides a tabular format the data are also presented as graphs and maps, which users can customize to their personal preferences. In its first two months the number of unique visitors doubled from around 15 to 30 thousand. Taking into consideration that the TGM always starts up with a Table view of the data, first web traffic statistics show that graph and map consultations sum up to a total of approximately 50%-80% of the total number of Table views.

10. From a technical point of view the introduction of the TGM is a relatively small step. In terms of user approach, however, the new visualisation tool is fundamentally different from the previous service of just providing numbers and downloadable data files.

11. Nowadays the website is not only focussed on specialists downloading data and carrying out analysis in their own environment, but more and more on the general public for whom pure numbers are less meaningful. This is where visualisation comes into its own, making the numbers speak visually: visualisation as a tool of bringing data in closer reach of the users.

12. Graphical tools can be broken down into three categories: maps, graphs and dashboards, each having many sub-categories. The displaying functionality of these graphics greatly differs from application to application, ranging from simple static images to more advanced dynamic graphical presentations with instruments on board (e.g. animation panel and/or toolbox for customization). In that sense the scope of what can be understood as a graphical tool is quite broad and difficult to limit.

² Visits to the website are defined as being unique and identifiable, where a single visit covers all page hits and is assumed to have ended if no actions were performed within 30 minutes. The number of unique visitors (IP addresses) varies on a monthly basis between 250 and 350 thousand.

13. Static images have a tradition of adding value to written analysis in publications. From an editorial perspective they should be eye-catching and attracting reader's attention, carrying a strong and simple message which supports the surrounding text. This editorial aspect also applies to dynamic graphical tool displays. However, instead of a human editor, the graphical tool should be accompanied by adequate mechanisms for delivering interesting data and appealing visual formats relevant from the user perspective.

14. Next to image handling and customization the visualisation tool may also comprise the functionality for printing, saving and bookmarking. Furthermore, the tool may also include the functionality for re-directing users to related information. By cross-referencing at the object level, each (object of) visualisation becomes a node giving access to visualisation of related objects, i.e. related data sets, related publications and related meta-data. This allows for a content driven type of navigation in a similar fashion as interrelated news stories on the web.

15. Besides tools for visualising single data sets, there are others providing a graphical overview across data sets, displaying a number of indicators simultaneously. Dashboard tools, Gapminder and the Business Cycle Tracer are examples of such crosscutting graphical overviews. They allow users to compare graphically two or more indicators in a number of ways, in time and/or cross-sectional. To some extent these advanced graphical crosscutting tools are targeting informed user groups such as students and business and politics, as they require some understanding skills.

IV. ESS TASK FORCE ON VISUALISATION

16. Currently map applications are the most widespread visualization tool for illustrating statistics graphically on the Internet. In 2007 Eurostat conducted a small survey on the use of map visualization tools at the ESS' National Statistical Institutes (NSIs). The main outcome of the survey is:

- The majority of NSIs judge the use of maps for illustrating statistics as important to very important and most have map applications in operation and/or under development.
- Despite the importance attached to map visualization the number of page views/users for the majority of NSIs does not exceed 2000 per month³. Although there is a general tendency of satisfied users, some NSIs mentioned a decrease in page views after the introduction of the application.
- The map applications can be broken down into two categories: specific software products for displaying maps such as ArcIMS, PX-Map or GeoClip, and tailor made applications based on, for instance SVG, Flash or the more recent Silverlight. There is quite a variety of map tools⁴ used across the countries. The functionality and statistical domains covered also varies widely. Some NSIs indicate that they intend to merge various existing map practices into a single approach and to extend also the coverage of (regional) statistical domains. Some of the problems that NSIs encounter are: missing automatic connection to the database, relatively high development costs, and a variety of technical solutions.

³ One of the reasons for a relatively low rate of usage could be that many of the map applications not directly accompanied by related background information such as tables and related information as described in recommendation points 8, 15, 17, 21 and 23 of this section.

⁴ The various technical solutions for map applications comprise: fixed images approach (JPEG, PDF, PNG), SVG, GIS-IMS, ArcIMS, other Web Map Services, Flash, PX-Map(2) which is based on SVG and GeoClip which is based on Flash, Google Map/Earth, JAVA, JAVA Script). The two main difficulties related to SVG applications are: compatibility with (new) browsers, and user reluctance to download an SVG viewer. Higher compatibility goes at the expense of special SVG features. The problem with GIS-IMS application is that its functionality is hidden and needs training, and the difficulty to combine the needs of "non-technical" or generic internet users and technical users. It is furthermore missing interactivity of many maps, graphic output and of course there are the license costs. ArcIMS does not provide everything needed for map applications, most importantly classifications are missing (they have to be provided in a query). This adds complexity on the client side. NSIs have expressed positive experiences with Flash applications, esp. with the graphic quality and distribution of Flash since it became the quasi web standard for interactive graphics and is now installed on 98% of all computers worldwide. Mixed experiences with JAVA solutions were indicated: good data handling, but bad graphic output.

17. In addressing these issues and discussing the challenges of improving visualization tools in general the ESS has launched the initiative of setting up a Task Force on visualization tools. The Task Force's main objectives cover (a) information on visualisation tool projects, including technical and organisational aspects (b) purposes and achievements of such tools, (c) identification of statistical domains for which there could be a user need for visualisation tools, (d) organisation of co-operation and (e) the creation of recommendations on the design and implementation of visualisation tools.

18. In its first meeting the Task Force identified a wide range of potential statistics for visualisation tools based on a series of criteria which should apply in the selection of statistics to be considered:

- (i) The statistics should be strongly demanded by users. It is important that this should either cover a wide range or a particularly important group of users. Monitoring and analysis of search terms can be used for identification of domains which are highly demanded. In any case the target group of domains and users should be identified as precisely as possible.
- (ii) Ideally the visualisation tool covers a repetitive and longer term user requirement. Illustrations that show a singular effect and are not reused by users are less useful than tools well integrated into the general dissemination.
- (iii) Very often a creative idea is required. Challenging the creative power of others (internally and externally) could result in unexpected new ideas. Sometimes a graphical tool of specific domain can be re-used for other domains.
- (iv) There should be an element of playing and fun with a tool.
- (v) Users need to understand the graphics and what happens in the tool.
- (vi) A link to a personal situation can be very useful. Examples are the personal inflation indicator, the life-expectancy calculator or the neighbourhood statistics using Google Earth.
- (vii) Visualisation is good to highlight certain developments, cross-sectional differences, and trends. Where accuracy is required, figures arranged in the traditional table view are better.
- (viii) The integration of different presentations, e.g. maps and tables is essential, only maps are not sufficient. Graphics should be accompanied by their underlying data.

19. In addition to the criteria listed above, visualisation tools can rely in general on the same design principles that apply to websites, for instance:

- (ix) Compatibility with different OS and different browsers. Related to this for example is the choice of static images versus vector graphics, and the type of software used.
- (x) Supporting search engine and search results.
- (xi) Addressing WAI issues. For WAI purposes an accompanying tabular format display of the data is recommended.
- (xii) Organised from user perspective. Allow the users to be in control.
- (xiii) The data is displayed effectively. For example, the graphic should automatically re-size depending on the window and screen setting of the user.
- (xiv) Availability of additional functions to modify the display.
- (xv) Navigation is consistent and opens up all content.
- (xvi) Relevant information about product, service and supplier. Ability to get advice from the supplier

- (xvii) The presence of related products, services and information.
- (xviii) Well-integrated into the website. The tool should not duplicate functions which are already available in the website, but instead link to those pages (e.g. Terms of use, FAQ/Help and Contact).

20. The Task Force furthermore agreed that the objective in the ESS should be to exchange not only experience but also programs. Exchange of software and other products and information should be as open as possible. However, the origin of tools and the development efforts should be well recognised. The inventory of existing tools – for which Eurostat's CIRCA library will be used - will be the vehicle for the exchange. As a first example Eurostat has provided the TGM source code and documentation on functionalities, IT architecture, installation procedures, etc. For sharing purposes the Task Force in addition recommended:

- (xix) To isolate as much as possible the input format for which there is a good degree of normalisation (XML, SDMX) from the technical display tool.
- (xx) To optimize the use of configuration files allowing modification of lay-out, headers, language, etc.

21. Besides general designing principles and criteria for identifying interesting visualisation tools, some specific suggestions for creating more user interest in visualized data can be mentioned:

- (xxi) Put fresh, aggregated and most popular information upfront, i.e. latest news releases. Users are generally more interested in receiving fresh, aggregated and popular information first and then in the immediate background more detailed underlying data and related products (such as related data sets, meta-data, and publications).
- (xxii) Start with a default display which is meaningful from a user perspective. For example a graphic which immediately reveals how a country compares to others and the latest trends. Providing supplementary image examples (Galleries⁵) covering additional graphical views on the data could support this approach, and are especially useful for users less acquainted with the tool.
- (xxiii) Users should be able to customize the graphic and to fine-tune the selection of data in an understandable way. Functions which are too advanced for basic users should be "hidden" under a button. One of the options for making graphics more personal is to allow users to insert their comment⁶ into the graph or to add a picture as wall paper. It is highly recommended to include a demo/tutorial.
- (xxiv) Facilitate the user in printing, saving, bookmarking and sharing their (customized) product. Standard report output (tiny publication) could also be envisaged, covering not only the selected/customized graph, but also related product information. Rolling out *sharing* concepts in an official statistical website is generally speaking less straightforward than for non-official data visualisation sites⁷, for instance as regards data upload by end-users, blogs and topic hubs.
- (xxv) Offer special features for frequent users for personalising the interface settings and for handling personalised traffic (e.g. My default display, My data sets, My data dimensions, My download format, My batch, notification). Another special feature is the instantaneous opening of visualized data in a user defined spread sheet system (pc or web). This indirectly enriches the

⁵ Galleries can range from simple window dressing (containing just a few predefined graphical images) to image libraries with search and filter functionalities such as "Most viewed", "Time range", etc.

⁶ Adding user comment i.e. analysis **into** the graph is based on a different concept than the sharing of comments on a graph between users.

⁷ Examples of some non-official data visualisation web sites: Swivel, ManyEyes, StatCrunch, Data360 and Gapminder. Web spreadsheets such as EditGrid are evolving in a similar direction, although they started from a different position.

functionality and usability of the visualisation tool. The tool can also offer widgets⁸ of the graphics which web/blog creators can copy/paste into their web (publication) page.

- (xxvi) Visualisation tools promote themselves by supporting (on-line) courses, for example on economy, and population and social conditions. For this purpose dissemination multipliers such as YouTube can also be considered. Students are one of the major user groups of statistical websites and can be regarded as future key users of statistics.
- (xxvii) Enhancing usability and visibility through an increased interconnectivity of the various statistical websites.

V. CONCLUSIONS

22. With graphical tools on the web the data are becoming more visible and more accessible to the general public than before. The increases in the number of data consultations clearly indicate the huge interest and first experience with Eurostat's data visualisation tool shows that graphs and maps are highly consulted.

23. Visualisation on the web of official statistics is relatively new and the ESS has recently initiated a Task Force on this subject. A number of recommendations and suggestions have been described in this paper.

24. Given the various initiatives undertaken for data visualisation and its relevancy for making statistics more interesting for the general users, developments could benefit from

- Consolidation of existing (best) practices on design and software;
- Sharing of (parts of) applications;
- Empowering creativity for new visualisation concepts. This is not only a matter of new designs; it is also connected to the identification of user needs which constitute the purpose and drive of creative data visualisations.
- Anticipating web trends⁹, from the IT perspective as well as conceptually (for instance as regards the sharing of non-official data and commenting; borderless mobility of content);
- Enlarging visibility (e.g. education, on-line courses; improving web interconnectivity between statistical websites; go where the public goes).

⁸ The widgets refer - according to Wikipedia - to "third party items that can be embedded in a frame" and are broken down in web widgets ("... anything that can be embedded within a page of HTML...adds some content to that page that is not static.") and desktop widgets ("... downloadable interactive virtual tools that provide services such as showing the user the latest news, the current weather.....a map program,...").

⁹ In the client-server relation the traditional approach was to put many demands on the web server for the delivery of new static html to the browser (client) with the disadvantage that after every http request the browser had to wait for a full page reply of the server. Since then a number of factors have boosted user experience, such as (a) graphical visualisation running under the browser (natively or as plug-in), (b) improved client-server logic via Rich Internet Applications (e.g. AJAX), (c) and increased bandwidth. It created some of the conditions for worldwide sharing of multi-media objects (YouTube, Flickr, etc.). One of today's challenges is to reduce (the maintenance regarding) browser incompatibility and OS dependency, making today's applications preferably CORE/A (Create Once, Run Everywhere/Anytime). Closely related to this are recent initiatives (Adobe's Integrated Runtime and Mozilla's Prism) to "re-package" web type apps into desktop apps running without browser and with the advantages of the local file system. From a user point of view the apps could become hybrid, i.e. in a web mode when on-line and in desktop mode when switched off from the Internet.