Dynamic graphics: Turning key indicators into knowledge

Enrico Giovannini

OECD Statistics Directorate

2 rue André Pascal

75016 Paris, France

E-mail: enrico.giovannini@oecd.org

Russell Penlington

OECD Statistics Directorate

E-mail: russell.penlington@oecd.org

Lars Thygesen

OECD Statistics Directorate

E-mail: lars.thygesen@oecd.org

1. Introduction

Information plays a vital role in economic and political processes. The increase in available information, coupled with advancements in Information and Communication Technologies (ICT) has changed the ways in which markets and societies work. The Internet, as well initiatives aimed to build information repositories through the collaboration of individuals, such as wikis and blogs, make more information available to more people than ever before. And so the ideal of a "fully informed decision maker" could be a reality. Unfortunately, this is far from the case: on the contrary, many people are overwhelmed and confused by the amount of information available. Gathering information has never been easier. Selecting the "right" information and turning it into knowledge is perhaps harder than ever.

While there is broad consensus on the need for high-quality statistics for decision making among policy makers and the business community, much less importance is given to the data needed by citizens to make their individual decisions. This is becoming increasingly important. People today are asked to make more decisions than ever before - decisions that can be vital to their long-term wellbeing (for example, choosing pension schemes, schools, or even what food to eat and how much, etc.).

Several software houses and large ICT companies are developing new tools to manage and disseminate statistics (especially indicators) to both specialised users and the public. Statistical offices and other research institutes are improving their dissemination and communication. Finally, civil society organisations have developed interesting tools to advocate their activities, to statistically monitor the implementation of official commitments by policy authorities and to foster facts-based democratic dialogue among citizens.

The OECD is experimenting new ways to disseminate its statistics using innovative visualization tools. This paper presents the main directions of work and some first results of the use of dynamic graphics.

2. Statistical information and citizens' knowledge

Economic and political theory affirms the importance that better information plays in increasing social welfare. But, information alone is not enough: it needs to be turned into knowledge and hence concrete behaviour. For information to become knowledge one must process, distil, digest and internalise available information and transform it into an idea or principle, subsequently used to take decisions.

If the production and dissemination of statistics is a service activity, its value, according to the "System of National Accounts", depends on the change it induces in the consumer's condition. For statistics, the change could be measured by the consumer's knowledge: after having looked at some statistical output the consumer should know something that they did not know beforehand. Unfortunately, evidence shows that citizens know very little about the key facts concerning the society in which they live. Therefore, for statisticians it is not enough to disseminate data and metadata to experts: statistics have also to be effectively communicated to the general public in order to help citizens to make better decisions and improve the democratic functioning of our societies. To do that all tools have to be exploited, especially those able to capture the attention of the final users and help them to get a sense out of the figures produced using statistical methodologies[1].

New technologies allow exploring new ways of communicating statistics. OECD decided to invest in this direction, experimenting new solutions, as well as new approaches to stimulate discussions among users about countries' economic, social and environmental conditions as described by statistical figures.

3. The Rome seminar on dynamic graphics

In order to scan the market for best practice regarding dynamic graphics, OECD arranged in cooperation with ISTAT (the Italian National Statistical Office) a seminar on 5-6 March 2007 in Rome (see http://www.oecd.org/oecdworldforum/graphics) The seminar was successful in allowing OECD to choose the most suitable solutions for further development on the dissemination layer of the OECD Statistical Information System. The seminar united a mixture of audiences with quite different cultures: statisticians who were keen to present their content, analysts and academia focusing on the messages themselves, and technical people mostly interested in what is inside the machine. This proved to be an excellent mix, fostering discussions and giving new insights.

In the Rome seminar, two types of graphs were showcased:

(a) General purpose tools, aiming to show almost any kind of statistics, e.g. as a front end to a data warehouse with many different topics;

(b) Special purpose tools, aiming to show a graph that is especially suited for just one kind of question or topic.

The special purpose tools can be more effective in the sense that they are adapted to the messages that the "sender" wants to get across. They can also give a more direct editing role to the sender, thus giving more power to the message. The general purpose tools may be much more widely applied, as any user can easily invoke the procedures to create them, but are prone to misuse or misunderstanding of

---

[1] When we look at the weather forecasts, what we see are symbols of "sun", "clouds", "rain", etc. Actually, what is behind these symbols is a probability associated to a particular weather condition, but the figures are invisible to us.

statistics. Then the user can correlate any data with one another, there is a risk of finding non-sense correlations.

Just as different techniques may be applied according to the problem or story that should be conveyed, there are differences individual preferences regarding the form of the presentation. While some may find the Bubble chart very convincing and easy to understand, others may think that they have learnt nothing from it, they would want a dynamic map, etc.

It is envisaged that special purpose and edited graphs may be presented effectively with accompanying videos, showing and commenting on the story.

4. Key directions for the OECD work

The OECD is presently developing a new dissemination policy for statistics, encompassing all the different outlets of statistics that have hitherto been split and incoherent. The aims are to enhance visibility and use of OECD statistics among existing and new audiences. The development of dynamic graphics is an element in this work. The new policy builds on a number of principles or directions.

4.1 Availability of data

Firstly, all statistical data and metadata have to be accessible in one place, the OECD statistical data warehouse OECD.Stat that has been built over the past three years. Whether users are actually allowed to get to the data will be decided by access rules. Access to the data warehouse will follow best international practices. Once data have been found, it can be presented in different forms, and one of them will be dynamic graphics, as described in 4.3. below.

A key point is that all statistical data and metadata need to be made available for easy reuse and reinterpretation by others, including the Web 2.0 community. Therefore, OECD will offer open and standardised access to data and metadata in the data warehouse. This implies general purpose web services from the database, using the SDMX-ML standard.

4.2 The use of new technologies

The ability to easily convey messages and highlight pertinent issues using statistics depends largely on the capabilities of technological solutions employed. It is well recognised that the most efficient platform to disseminate statistical data and graphical representations of statistical data to a wide audience is the web. Internet standards have made it difficult in the past to render dynamic graphical content to users due to limitations for cross platform support within browser software and limitations within the World Wide Web Consortium (W3C) Hypertext Mark-up Language (HTML) specifications. Efforts have been made to overcome these limitations by offering third party plug-ins for Internet browser software that allow rich content to be rendered to users outside of the HTML standards.

While such plug-ins are free to end users, some have been based on open source standards while others have required proprietary software in order for development to be made. There is, however, an evolution towards open source standards with optional development environments simply to accelerate deployment. A classic example of this movement is the Scalar Vector Graphics (SVG) versus Flash story. SVG is an open source W3C standard for rendering XML data graphically online. SVG has therefore been

employed vastly on the web to render graphical visualisations of data. To do so, however, requires considerable development as programming must be accomplished at a very granular level. Flash, on the other hand, requires less development resources as the graphical rendering is handled in more of a component style by the internet browser plug-in (Flash Player) required to be installed on the user's computer. The disadvantage of classic Flash applications is that it requires third party software for development.

SVG is now becoming obsolete as software vendors are dropping support for it in favour of their own technologies. Adobe (who provides Flash development environments) has traditionally provided an SVG plug-in for Microsoft's Internet Explorer browser. Other browsers such as Mozilla Firefox natively support SVG without a requirement for a plug-in. Adobe will no longer support the SVG plug-in for Internet Explorer and Microsoft are developing an alternative proprietary technology.

Adobe has developed a Rapid Application Development (RAD) environment called Flex that generates rich Flash content for displaying graphical content online. While Adobe does provide an Interface Development Environment (IDE) at a cost, the overall technology is free and open source. Flash provides a much richer user experience and is easier to develop than SVG, so this evolution to open source and cross platform standards that require a free plug-in for the user makes sense.

The OECD is piloting the use of Adobe Flex to display statistical data graphically online. In order to ensure the portability of developments to the greater statistical community, this development is based on content in the Statistical Data and Metadata Exchange (SDMX) ISO standard. The OECD is working in parallel with the European Central Bank (ECB) to create a Flex application that can interrogate SDMX data structure definitions and allow the user to view SDMX-ML data graphically and in tabular format. Focus has been made on two of the more common formats of the SDMX standard: generic and compact data messages. The ECB employs the compact format of the SDMX standard to provide data online for users and the OECD provide a web service for statistical data in the generic format of the SDMX standard. Adobe has expressed interest in this initiative and is participating in development.

The development that the OECD and ECB are undertaking based on the open source Flex technology and the SDMX ISO standard will be made available to the greater statistical community in order to facilitate the visualisation of statistics in rich dynamic graphical formats. Such an effort can also help to promote the employment of the SDMX standard to disseminate and share statistics which is an important initiative to provide consistency for statistics and resource efficiencies.

5. New approaches to communicate statistics: some recent OECD experiences

Since January 2007 the OECD has launched several initiatives to explore new ways of presenting its data. What follows, highlights some of the concrete developments that are taking place.

5.1 The OECD Factbook in Swivel

After the end of the dot-com bubble in the fall of 2001, some people concluded that the web was over-hyped. More recently, the concept of "Web 2.0" began to be used to indicate a new approach to web developments and over the last two years the term "Web 2.0" has clearly taken hold, with more than 9.5 million citations in Google. But what are the main characteristics of the "Web 2.0"?

The first one is that the "web is a platform" where users benefit of web services. For example, Google

began its life as a native web application, never sold or packaged, but delivered as a service, with customers paying, directly or indirectly, for the use of that service. Much like a phone call, which happens not just on the phones at either end of the call, but on the network in between, Google happens in the space between browser and search engine and destination content server, as an enabler or middleman between the user and his or her online experience. By alluding to the version-numbers that commonly designate software upgrades, the phrase "Web 2.0" hints at an improved form of the World Wide Web; advocates suggest that technologies such as weblogs, social bookmarking, wikis, podcasts, RSS feeds (and other forms of many-to-many publishing), social software, Web APIs, Web standards and online Web services imply a significant change in web usage. As used by its proponents, the phrase "Web 2.0" can also refer to the transition of web sites from isolated information silos to sources of content and functionality, thus becoming computing platforms serving web applications to end-users. A social phenomenon embracing an approach to generating and distributing Web content itself, characterized by open communication, decentralization of authority, freedom to share and re-use, and "the market as a conversation".

To explore these new opportunities, the OECD decided to make available on a Web 2.0 platform the data published in its "Factbook", a selection of more than 200 economic, social and environmental indicators. After some market exploration, the OECD decided to upload the Factbook data and metadata on www.swivel.com, a "data exploration web site for curious people", as they define themselves. More precisely, Swivel is a web site for uploading, exploring and sharing data. Individuals and organizations worldwide can easily upload their data to Swivel, compare it with other data and share their insights via email, Web sites and blogs. Launched in December 2006, more than 1.6 million charts have been created in a few months.

Loading OECD core data on Swivel, which created a special label "Official Source" to distinguish data uploaded by the OECD and by individuals, helps the Organisation to extend the discussion of its valuable economic facts and figures beyond its traditional audiences, helping to meet its goal of building knowledge based on fact. OECD's goal is also to encourage a dialogue among those who use our data, so we can learn more about the data they need, raise the visibility for our core data among a new audience and see how they react, communicate and share their thoughts in a creative and open environment.

5.2 The OECD Factbook in Trendalyzer

In the context of the second OECD World Forum on "Statistics, Knowledge and Policy" (Istanbul, 27-30 June 2007, see www.oecd.org/oecdworldforum), an exhibition on innovative tools to help people to turn information into knowledge will be organised. Stands will be made available to demonstrate innovative software and tools to disseminate data in a way to facilitate knowledge; and the world's best indicators systems will also be showcased. Attendees will be asked to express their views on the various proposals and, on the basis of attendees' preferences "World Forum Awards" for the most innovative solutions to transform statistics into people's knowledge will be assigned during a session, when the winners will have the opportunity to present their solutions during the plenary session aimed to discuss how to transform statistical information into knowledge.

Actually, the second World Forum is designed to allow participants to make a true "knowledge experience". In co-operation with the Gapminder Foundation (www.gapminder.com), the OECD "Factbook" data will be uploaded on the Trendalyzer software, and used to display economic, social and environmental

phenomena throughout the days of Forum[2]. During the Forum days plasma screens placed in coffee breaks spaces, in the exhibition area, etc. will be used to present Trendalyzer animations based on OECD data, to help participants to focus on how our societies are progressing or worsening.

5.3 Key indicators to evaluate countries' performance: a new tool based on "spider charts"

One of the key issues for organizations that produce statistical indicators concerning countries' performances is how to represent, in a synthetic way, the relative position of each country vis-à-vis the others or relevant groups of countries (OECD totals, EU totals, etc.). The problem becomes even more complicated if the presentation needs to be done taking time into consideration.

The OECD has tried to develop a possible solution to this problem, developing "dynamic country profiles" based on 32 indicators (derived from the Factbook), divided in four groups (two concerning economic dimensions, two social dimensions), represented trough dynamic "spider charts" (or "radar charts"). Each chart presents 8 indicators concerning the selected country and the OECD total/average. The user can both select a particular year or the "animated presentation". In the latter case, 5-year moving averages are used to show how the country's situation evolved over time in comparison with the OECD total.

To make the charts more meaningful some manipulations of the original data were necessary. In fact, each indicator has its own scale and therefore a preliminary standardization was necessary. Moreover, to
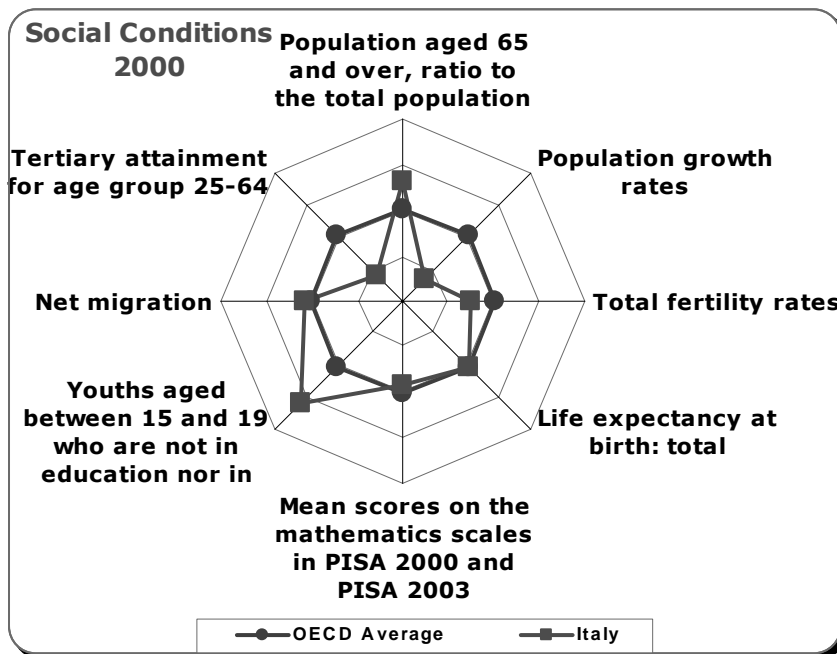

Figure: Example of Spider Chart

---

[2] Actually, the first OECD World Forum (held in Palermo, in November 2004) was one of the first opportunities for the international community to know about the Gapminder software, which changed the way in which many statisticians look at the communication of statistics.

make the charts more easily interpretable, some indicators are presented using an inverted scale[3]: in this way, the ideal situation for a "virtuous" country is to have all values of indicators higher ("external") than those referring to OECD totals, although in some cases the interpretation of indicators is not so straightforward.

The new system for animated "spider charts" will be made available on the OECD web site also to be downloaded and used off-line. Of course, the downloadable file will contain all data and metadata for the 32 selected indicators. It will also be used to produce colour charts to be used for presentations of specific country's conditions by OECD officials.

6. Conclusions

Statisticians have a lot of work to do to increase the actual use of statistics by the general public. This is a "must" in the information age and in a situation in which more and more data sources are potentially available. Help users to better understand the key messages emerging from statistical figures is part of the today statisticians' work and cannot simply be delegated to classical media channels.

New technologies allow the adoption of innovative approaches to communication of statistics, as well as the involvement of users in group discussions. The OECD is experimenting some of these approaches as part of its new dissemination strategy. For a "knowledge builder" institution like the OECD the use of dynamic graphics and other Web 2.0 tools to disseminate statistics is fundamental to further strengthen its role in the global information market.

RÉSUMÉ (ABSTRACT)

L'OCDE s'est engagée à promouvoir la prise de décisions basée sur l'information dans les différents aspects de la vie : élaboration de politiques, décisions économiques et vie privée des citoyens. L'OCDE est un fournisseur des statistiques comparables de haute qualité qui se prêtent à la prise de décision et l'analyse des politiques internationales. Cependant, les décideurs doivent d'abord prendre conscience et comprendre les messages enfouis dans ces données.

Pour beaucoup de personnes, et particulièrement les plus jeunes, les statistiques sont rébarbatives et difficiles à comprendre. De ce fait, elles ne peuvent se transformer en connaissance. Cet article s'attache à présenter le défi qui consiste à faire en sorte que les utilisateurs potentiels considèrent les statistiques comme intéressantes et utiles pour se forger une opinion et prendre des décisions Il propose d'utiliser des méthodes graphiques dynamiques pour rendre les indicateurs attractifs et permettre de découvrir les messages qu'ils cachent. Il propose en outre une plus grande utilisation d'approches participatives sur l'Internet, encourageant les utilisateurs à expérimenter avec leurs propres analyses et à discuter des résultats et de leurs implications.

---

[3] This is made immediately visible to the user as the label of the indicator with an inverted scale has a different colour from the others.