WIKIS, DYNAMIC CHARTS, VIDEOS AND OTHER INNOVATIVE TOOLS TO TRANSFORM STATISTICS INTO KNOWLEDGE

Enrico Giovannini
President of the Italian Statistical Office (Istat), Rome, Italy
Enrico.giovannini@istat.it

The development of web 2.0 and other Information and Communication Technologies (ICT) are creating a revolution in the way information is produced and shared among different interest groups and individuals. Concepts like "collective intelligence", "crowdsourcing" and "prosumers" are at the basis of successful initiatives like Wikipedia, Innocentive, Facebook and other platforms used to develop both free and fee products and services widely appreciated, especially by new generations. How are these trends affecting the statistical world? Can "official" data providers continue to play their role just by introducing new ICT tools without changing their business model? The purpose of this paper is to discuss where the valued added of statistics comes from and the way in which information is spread in society. Some experiences about the use of innovative communication tools/approaches are also highlighted.

THE VALUE ADDED OF STATISTICS: WHERE DOES IT COME FROM?

Economic statisticians, and especially national accountants, have developed guidelines on how to measure the value added of each and every economic activity, but very little effort has been made on the measurement of the output and the value added associated to the work of national statistical offices (NSOs) and international organisations producing statistics. A survey carried out in 28 countries (See http://www.unece.org/stats/documents/ece/ces/bur/2008/25.e.pdf) indicates that the most frequently used output indicators include: number of publications (or number of releases); number of publication copies sent to subscribers; number of visits to the Internet page; number of indicators accessible in the Internet databases; number of tables viewed in the Internet databases; number of presentations at conferences and seminars; number of media quotations. Many NSOs also try to measure the quality of output with quantitative indicators (punctuality of releases, number of errors discovered in published information, revisions in statistical database, etc.) or user's satisfaction surveys.

Of course, all these measures are very important to monitor the implementation of the work programme and the usage of statistics. But can we really say that they are good measures of output and/or value added of official statistics? Looking at the statistical standards developed to measure economic activities, it emerges that according to:

- the International Standard Industry Classification (ISIC Rev.1), the production of official statistics is a non-market service;
- the 1993 System of National Accounts, services are the result of a production activity that changes the conditions of the consuming units;
- Atkinson (2005), "the output of the government sector should in principle be measured in a way that is adjusted for quality, taking into account of the attributable incremental contribution of the service to the outcome".

But what should the final outcome of official statistics be, considering what the SNA says? As reported by Wikipedia, the Oxford English Dictionary defines "knowledge" variously as: (i) expertise, and skills acquired by a person through experience or education; the theoretical or practical understanding of a subject, (ii) what is known in a particular field or in total; facts and information or (iii) awareness or familiarity gained by experience of a fact or situation. Therefore, "knowledge" seems to be the answer: knowledge of economic, social and environmental phenomena. If a person knows nothing about a particular issue and looks at relevant statistics, should s/he not become more knowledgeable (to a certain extent) about that subject?

Therefore, we could conclude, that the value added of official statistics (VAS) is linked to what the actual (not the potential) users know about the facts that are relevant to them in making their decisions. In short, from a collective point of view the production of official statistics can change according to two factors: the size of the audience (i.e. the number of people who know

official statistics, N); the quantity of official statistics (QS_i) actually included in the information sets relevant to each individual's decisions

$$VAS = \sum_{i=1}^{N} (N * QS_i) - CS$$

while CS represents the cost of producing statistics. If only a small group of people is aware of official statistics, it is less likely that society will use them to take decisions. On the other hand, if everybody knows about official figures, but do not actually use them when making decisions, their value added will be minimal.

Globalisation, information society and political reforms (that require individuals to take decisions that in the past were taken by the government – pensions, education, etc.) are making N bigger than ever. On the other hand, QS can depend on several factors, such as:

- the total amount of official statistics that reaches a generic user (QSR_i). This amount depends on two elements:

$$QSR_i = QSA * MF_i$$

  where QSA represents the total statistical information produced by the official source and the role played by media (MF) for each individual, which can emphasise or reduce the actual amount of information communicated to the generic user;
- the relevance of the official statistics communicated to the user (RS_i);
- the trust that individuals have in official statistics (TS_i);
- the individuals' "numeracy" (i.e. the ability to reason with numbers and other mathematical concepts, NL_i).

We could then write the following expression:

$$VAS = \sum_{i=1}^{N} \{[(QSA * MF_i) * RS_i * TS_i * NL_i]\} - CS$$

Of course, it is extremely difficult to quantify the different elements that enter in the equation. However, some sparse evidence exists. For example, as described in Giovannini (2007):

- 69% of European citizens believe that it is necessary to know key economic data (such as GDP, unemployment rate, inflation rate, etc.), but 53% of European citizens have no idea of what the GDP growth rate is in their country and only 8% know the correct figure (similar figures have been obtained by Curtin for the United States;
- 45% of Europeans tend not to trust official statistics, while 46% tend to trust them;
- in the United States, the most common source of information on official figures is TV (78%), followed by newspapers (58%), Internet (37%), radio (34%), family/working networks (34%) and magazines (14%). The five main TV networks quite frequently report data on the unemployment rate (83% of cases on average), but much less frequently data on GDP growth (46%) or inflation rate (35%). Looking at the 27 most popular newspapers, on average they cover just 39% of the official reports on GDP, 53% of those concerning CPI and 52% of those announcing the official unemployment rate;
- finally, when disseminating US economic data, the Associated Press and United Press International (the most popular wire services) typically do not mention specific source agencies in their releases. This approach has a clear impact on the "brand name" of the source: 23% of Americans have never heard of official unemployment data or the source agency; the comparable figures are 34% for CPI and 40% for GDP.

This review underlines three key points for the following discussion: first, the way in which statistics is used/perceived by users (especially citizens) depends on several factors and some of them are not under the control of the original source; second, in several countries the situation is far from being satisfactory in terms of trust in and communication of official statistics;

third, statisticians have to address these issues (measurement of their output and value added, relationships with media and final users, brand image, etc.) very seriously, especially if they wish to respond to the challenges coming from the web 2.0 revolution.

THE DISSEMINATION OF INFORMATION AND THE ROLE OF WEB 2.0

As Einstein said, "information is not knowledge": therefore, what people know must not be confused with the amount of information they receive every day and absorb from the most disparate sources. Instead, knowledge (which ultimately represents the value added of statistics) refers to a complex and dynamic process involving cognitive mechanisms. Several models have been developed to explain how these mechanisms work, and one which is particularly relevant to this discussion is the model based on the so-called "epidemiologic" approach (see Sperber, 1996). Originally developed for cognition and culture, the concept of epidemiology has been increasingly applied to the study of a wide range of phenomena and recently, economists have also begun to refer loosely to epidemiological processes for economic modelling. In a nutshell, this approach states that information is spread like a virus across society. At the beginning only a few people catch it, but then each "infected" person transmits it to others, and so on. However, every time there is a transmission, the information changes a little, as viruses do.

If information is spread across society as a virus, which evolves with every passage, it would be fundamental for statistical data providers to reach as many people as possible at the beginning of the chain, to "vaccinate" them against the "ignorance disease". But to do that, they have to: disseminate information relevant to people, present information in a way that people can relate it to their own interests, using language/tools coherent with those used by people in other contexts.

Of course, data providers are aware of these problems and have heavily invested resources to improve their communication tools. But new ICT tools and the success of the web 2.0 are also profoundly changing the way in which people, especially new generations, look for and find data, or build their knowledge. The use of web 2.0 transforms the "consumer" of a particular information/service provided via the Internet into a "prosumer", i.e. a person that is simultaneously a consumer and a producer of the information/service. Of course, Wikipedia is the most popular example of this approach, but there are many other platforms that use "collective intelligence" (defined as "a form of intelligence that emerges from collaboration and competition by many individuals") to develop innovative services. This change has deep implications for the world of official statistics, which is becoming more aware of the need to exploit the opportunities offered by web 2.0, but it can be easily said that prudence is dominating. The next sections provide some examples of what has been recently done by national statistical offices and some international organisations, starting with the use of visualization tools.

NEW VISUALISATION TOOLS

This is an area where, over the last five years, official statisticians have made important changes. Since Hans Rosling gave his presentation of Trendalyzer at the OECD World Forum on "Statistics, Knowledge and Policy" in 2004, shocking the large audience who attended the Forum and opening people's eyes about the opportunities provided by dynamic visualization, statistical offices and international organizations have paid more attention than ever to this area of work. At the same time, private ICT companies and research centres, as well as designers and media, have invested a lot to improve the way in which data are presented and visualized. For example:

- OECD has been very active in both organizing international events on new ways to present statistics (www.oecd.org/progress) and using new tools to make its data available to the public. For example, the *OECD Factbook* has been made available using Trendalyzer, OECD eXplorer, other dynamic graphics tools, mobile devices, etc.;
- IMF has made a wide range of economic data available using IMF Data Mapper (http://www.imf.org/external/datamapper/index.php), a dynamic visualization tool;
- The European Central Bank has developed a dynamic "Inflation dashboard" using Flash;

- The World Bank has recently made available the World Development Indicators via a dynamic visualization tool developed using Prognoz (http://devdata.worldbank.org/DataVisualizer);
- Some statistical offices (especially in Netherlands and United Kingdom) have created units to develop visualization tools (for example the "Business clock", also used by Eurostat) and implemented new ways to present statistics to the wide public.

At the same time, the number of tools available for visualization has experienced an unprecedented growth. A recent review (http://www.insideria.com/2009/12/28-rich-data-visualization-too.html) listed 28 different visualization tools and each of them is able to provide a variety of visualizations unavailable until few years ago. Moreover, Swivel.com and ManyEyes.com provide not only visualization tools, but also platforms to share charts. The development of such a variety of tools has also promoted an emerging community of experts in this field. Besides the international exhibitions organised by the OECD since 2007, or the meetings attended by software developers working in national statistical offices, international organizations and research institutes, several websites have been established to foster dialogue and collaboration on these issues. For example, FlowingData.com "explores how designers, statisticians, and computer scientists are using data to better understand ourselves - mainly through data visualization".

Given the variety of software available and the user-friendliness of these tools, one could expect that the websites of statistical offices and international organisations are full of dynamic charts. Unfortunately, this is not the case. Besides some notable exceptions, like those mentioned above, most of NSOs' websites are still using quite standard charts. Why? There are several answers to this question and we will try to discuss some of them in the concluding section, after discussing other recent developments in this area, namely blogs, videos and wikis.

BLOGS

Statisticians are well aware of the success of blogs, as well as of their power to engage people in lively discussions. So, one could expect that, given the lack of confidence that some people have in official statistics, they look at blogs as a way to overcome the distance between producers and users of statistics. Quite clearly, this is not the case. And there are good reasons for that. To discuss them, we will make reference to the OECD case.

When in 2008, the OECD Statistics Directorate decided to restructure its Statistics Portal, it was decided to include an area, called "Spotlight", to post every one or two weeks one chart and some text about a "hot topic", with a blogging area. The result was disappointing: very few comments were posted, no matter what topic was chosen and a similar results was obtained by the IMF, who tried to establish blogs to discuss key speeches given by its top managers.

Of course, we tried to understand the reason for such a failure, especially considering that the new OECD Statistics Portal was regarded as excellent by users and the number of hits increased a lot. Our answer was rather straightforward: besides the fact that it was quite strange to find a blog on a Statistics Portal (normally visited to find data) and not on the pages managed by policy analysts, we found out that people did not consider the OECD website a place for discussion, but only a tool to find and download interesting information or data, and if necessary to use them to develop their analytical work and then take position on their own blogs or on sites dedicated to a particular topic. Therefore, the main lesson we learned was that it is not enough to use web 2.0 tools to create a community and make it work To do that, potential participants have to see the institution hosting the blog as open and transparent, ready to engage itself in a frank discussion with external people.

As things stand, it can be easily understood why statistical offices do not use blogs. In fact, given their role, they would be in trouble in discussing the substantive issues raised by the data they publish, while discussing only statistical issues (sample techniques, estimation methods, etc.) in a blog does not seem to be a good idea. However, blogging could be useful to address issues raised by media or by the public that can influence the trust in statistics. Of course, this should be done rather carefully, but the idea should not be discarded without a thorough evaluation. Let's think, for example, about the criticisms addressed to some European statistical offices after the euro

changeover, when large part of the population (and media) criticized official inflation data, considering them clearly underestimated. What if the chief statisticians of those countries had opened a blog to have an open discussion with citizens, civil society organizations and media on these issues? Would have this helped explain the correctness of inflation data, thus avoiding the loss of confidence in official statistics produced by people's criticisms? Of course, it is very difficult to provide a sensible answer. But it is true that, in the web 2.0 world, the channels through which legitimacy and authority are built are quite different from the past and official statisticians should be open to test different and innovative ways to keep, or increase, visibility and respect vis-à-vis their societies.

VIDEOS

Also in this area we have to quote Hans Rosling, who developed what he called "Gapcasts" (www.gapminder.org), i.e. short videos to discuss a particular issue (infant mortality, CO2 emissions, etc.) using Chroma Keying. This technique, commonly used for weather forecast broadcasts wherein the presenter appears to be standing in front of a map, was used by Rosling to develop videos where he provided, in a very lively way, comments on the dynamic charts generated using Trendalyzer, then posted them on YouTube.

Quite impressed by this initiative the OECD tried to implement it in a more systematic way, taking the following actions:

- test the use of Chroma Keying, to identify challenges from "technical" and "human" points of view. As usual, the technical side is the easiest one, while playing the role of "storyteller" requires specific skills and training;
- identify some possible stories to be presented. This is not so easy to do because Trendalyzer is particularly effective when long time series are available (twenty years at least);
- contact international media networks (Bloomberg, BBC, Reuters, etc.) to reach an agreement for the production and the dissemination of the videos.

Besides the "human" and statistical challenges, it was impossible to find a media partner to implement the idea. After the initial enthusiasm, technical difficulties emerged (interoperability between Trendalyzer and media's software and video systems), as well as political challenges (who is in charge of choosing the topic, preparing the story, telling the story, etc.), so nothing happened in practice.

An interesting development of Rosling's approach was made during a summer course organized in 2008 by the Kessler Foundation (http://webvalley.fbk.eu) for a small group of highly IT skilled teenagers, where the idea of Gapcasts was presented and discussed. At the end of the training course a video was produced about the evolution of life expectancy in Uganda. Using a touchscreen monitor, a teenager presented a story where the classical dynamic charts obtained using Trendalyzer were integrated with pictures of Ugandan leaders (Obote, Amin, etc.) and of main events (wars, revolutions, etc.) that explained changes in life expectancy. These images added a great deal of realism to the story presented by the speaker.

What can we conclude from these experiences? First, the use of videos for storytelling has to be seen as a key tool for the communication of statistics. Therefore, statistical offices should carefully study this possibility, looking at both content and organisational challenges. Second, it is easier for a public institution to directly implement this approach, but if the aim is to reach a wide public an agreement with media networks becomes indispensable, although the implementation of the initiative could be much more complicated.

WIKIS

Wikis can be used in different ways: for example, to develop a product fostering collaboration among authorized people before it is made available to external users, or to use a publicly accessible platform to develop a product through the contribution of whoever is willing to do that. Besides the use of wikis to foster the dialogue between methodologists and analysts working in European statistical offices, Eurostat used the first model to develop "Statistics explained", a site (developed using Mediawiki, the software also used for Wikipedia) presenting a

sort of encyclopedia of European statistics, completed by a statistical glossary (clarifying all terms used), several links to further information and the very latest data and metadata (http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Main_Page).

The second model has been used by the OECD to develop "Wikigender", launched in March 2008 using Mediawiki, with the aim to share and exchange information on the situation of men and women around the world (www.wikigender.org). It welcomes an active participation from users who can contribute to the content of the website by posting comments, editing articles or creating new entries into this knowledge database. The statistical part of Wikigender is based on a series of links to external databases, including the OECD ones. Moreover, for each country a subset of indicators is presented in tabular and graphical forms.

Wikigender was also developed as a preliminary step towards the implementation of "Wikiprogress", launched at the end of 2009. Wikiprogress has been designed to be the catalyst of initiatives existing around the world on the measurement of societal progress, as well as their use for raising awareness amongst stakeholders, informing them on key economic, social and environmental trends and allowing them to discuss relevant issues based on solid evidence. To achieve these objectives, Wikiprogress will be a multi-purpose website, with two main parts:

- a classical wiki (similar to the current Wikigender) where users can find, in addition to information about existing or new initiatives aimed at measuring progress around the world, contributions by those who run these initiatives and/or other people, as well as information about the measurement of particular phenomena;
- a "Statistical Wiki", where data and metadata can be shared, assessed, uploaded at the end of a "quality-assurance" process, and finally made available to users using state-of-the-art visualization technologies (such as OECD eXplorer).

Wikiprogress represents a highly innovative project, both from a technical and organisational point of view, with a lot of technical and governance challenges, but it could be the new world platform to host high-quality statistics produced by both public and private institutions, where people can access "certified" data on all domains and play with them applying the philosophy of "collective intelligence" to the development of statistics-based knowledge.

CONCLUSIONS

This review shows the great opportunities offered by new technologies to improve the communication of statistics and increase its value added. At the same time, it also shows some of the obstacles that make this process slow and complicated. Technical, organizational and cultural barriers can prevent official data providers from using available tools, thus negatively affecting their overall position in the "information market". To overcome these obstacles top managers of statistical offices and international organizations should: first, devote attention to these issues, understanding the cultural shift that web 2.0 is producing in the way in which information is collected, shared and used to build knowledge; second, establish organizational units devoted to carry on this work and to help the organization in making the necessary changes to take advantage of these innovations; third, foster the dialogue with private companies, research labs and media experts to anticipate future changes in technologies and cultural models, in order to innovate their organizations and bring statistics at the core of the information revolution that is happening no matter whether we like it or not.

REFERENCES

Atkinson A. (2005), *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts,* Palgrave Macmillan, New York.

Curtin, R. (2007). *What US Consumers Know About Economic Conditions*, paper presented at the OECD second World Forum on "Statistics, Knowledge and Policy", www.oecd.org/oecdworldforum.

Giovannini, E. (2007). *Is globalisation a threat to official statistics?*, paper presented at the DGINS Conference, September.

Sperber, D. (1996). *Explaining Culture. A Naturalistic Approach*. Oxford: Blackwell Publisher.