Committee for the Coordination of Statistical Activities

Conference on Data Quality for International Organizations
Rome, Italy, 7 and 8 July 2008

Session 3: Dissemination platforms to make data more accessible and interpretable

The role of communication in transforming statistics into knowledge

Enrico Giovannini (Chief Statistician of the OECD)

1. Introduction

The development of web 2.0 and other Information and Communication Technologies (ICT) are creating a revolution in the way in which information is produced and shared among different interest groups and individuals. The success of Internet platforms where communities create information through interactions provides evidence that the well consolidated roles of producers and users of information are radically changing. 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 in just introducing new ICT tools (web sites, visualisation tools, etc.) without changing their business model, or do they need to deeply re-think their classical role of information providers to evolve towards something else? What kind of approach do data providers need to develop to contribute to the functioning of a modern democracy in the "information age"? And how can communication strategies help in this respect?

In this paper, we will first discuss where the valued added of statistics comes from. Then we will deal with the way in which information is spread in society and how ICT tools are changing the paradigm of "societal knowledge building". Some experiences made by the OECD about the use of innovative communication tools/approaches will be highlighted in the fourth section. Finally, some conclusions will be presented.

2. 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 recent survey carried out on 28 countries[1] indicate 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? If we look at the statistical standards developed to measure economic activities, we find that:

- according to the International Standard Industry Classification (ISIC Rev.1), the production of official statistics is a non-market service[2];
- according to the 1993 System of National Accounts, services are the result of a production activity that changes the conditions of the consuming units[3];
- according to 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? "Knowledge" seems to be the answer: knowledge of economic, social and environmental

[1] See http://www.unece.org/stats/documents/ece/ces/bur/2008/25.e.pdf.

[2] It is part of Section L, Division 75 "Public Administration and Defence", Group 7511 "Administration of the State and the economic and social policy of the community", which includes "administration and operation of overall economic and social planning and statistical services at the various levels of government".

[3] In particular: "The changes that consumers of services engage the producers to bring about can take a variety of different forms such as: (a) changes in the condition of the consumer's goods: the producer works directly on goods owned by the consumer by transporting, cleaning, repairing or otherwise transforming them; (b) changes in the physical condition of persons: the producer transports the persons, provides them with accommodation, provides them with medical or surgical treatments, improves their appearance, etc.; (c) changes in the mental condition of persons: the producer provides education, information, advice, entertainment or similar services in a face to face manner".

phenomena[4]. 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?

We could conclude, therefore, 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. Therefore, from a collective point of view this value 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) actually included in the information sets relevant for each individual's decisions:

$$VAS = N * QS$$

If only a small group of people is aware of official statistics, the probability of society using them to make decisions is relatively small. On the other hand, if everybody knows about official figures, but individuals 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). This amount depends on two elements:

  $$QSR = QSA * MF$$

  where QSA represents the total statistical information produced by the official source and the role played by media (MF), 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);
- the trust that individuals have in official statistics (TS);
- the individuals' "numeracy" (i.e. the ability to reason with numbers and other mathematical concepts, NL).

[4] 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.

We could then write the following expression:

$$VAS = N * [(QSA * MF) * RS * TS * NL]$$

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 (2006):

- 69% of European citizens believe that it is necessary to know key economic data (such as GDP, unemployment rate, inflation rate, etc.)[5], but 53% of European citizens do not even have a vague idea of what the GDP growth rate is in their country and only 8% know the correct figure[6];
- 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[7];
- Finally, when disseminating US economic data, 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

[5] These data were collected in 2007 by the European Commission (Eurobarometer) at the OECD's request in preparation for the second OECD World Forum on "Statistics, Knowledge and Policy" (www.oecd.org/oecdworldforum).

[6] Similar figures have been obtained by Curtin (2007) for the United States.

[7] "If we presume that the 27 papers with the largest circulations all had access to the wire reports, the lack of complete coverage would be an active decision of the newspaper to not carry the report. It was likely to reflect a judgement about the newsworthiness of the latest figures given their subscribers' interests. There was a tendency for newspapers to more frequently report the latest official figures when it represented an unfavourable development, which may reflect the greater importance people place on the information content of 'bad' news" (Curtin, 2007)

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.

3. How is information disseminated?

We have seen that several obstacles can make the transmission of information difficult to potential users. Moreover, 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)[8]. In a nutshell, it states that information is spread like a virus in a 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.

In this context, three points require special attention:

- the amount of news released by the media plays a key role in affecting what people know;
- the quality of media and their way of presenting information can make a huge difference on people's capacity to grasp the sense of the what is communicated;
- the degree of exposure to the media is not sufficient for a person to be properly informed and to process the news, but the person's interest in the subject plays a key role in activating the cognitive mechanism.

[8] 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 particular, Carroll (2001 and 2002) has recently provided a new explanation of the way in which expectations that appeal to epidemiology are formed among people.

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. Unfortunately, this is not what statistical data providers normally do. Instead, they rely heavily on mass media, who in several cases do not help in spreading the correct information. So, is there an alternative?

4. The Web 2.0 revolution

Of course, data providers are aware of these problems and have heavily invested resources to improve their communication tools, especially the use of Internet. But new ICT tools and the success of Internet are also profoundly changing the way in which people, especially new generations, look for and find data. For example:

- according to Internet experts, 95% of those who use Google do not go beyond the first page of occurrences; and once they reach a particular site, a similar percentage of users do not click more than three times to find what they want: if after three clicks they have not found what they are looking for, they quit the site;
- the way in which "discovery metadata" are structured is fundamental to be placed in the first page of Google's results, but these metadata have nothing to do with the intrinsic quality of the information provided. Therefore, sources able to structure their "discovery metadata" well can appear higher than those which have better quality information but do not invest in this kind of metadata.

Everybody knows the most popular tools and success stories developed by the Internet community over the last few years. Maybe, less people are aware of the deep changes that the web 2.0 is producing in the way in which "collective knowledge" is generated today using "wikis", and how this is affecting the "digital native" generation's thinking[9]. Why is this so important for

[9] *Web 2.0* refers to a perceived second generation of Web-based communities and hosted services – such as social networking sites, wikis and folksonomies – which aim to facilitate collaboration and sharing by users. The main difference between the first and the second generation of Internet platforms is that the former are mainly "repositories of information", while the latter are "marketplaces" where people exchange and share information, meet people,

our discussion? The main reason is that this approach tends to transform the "consumer" of a particular information/service provided via Internet into a "prosumer", i.e. a person that is simultaneously a consumer and a producer of the information/service. Wikipedia is the most popular example of this approach, but there are many other platforms that use "collective intelligence" to develop innovative services.[10]

Of course, reliable statistics cannot be generated using "collective intelligence", but this does not mean that this approach does not have a huge impact on the way in which statistics are perceived or used. If, for example, an authoritative member of a "community" spreads the information that a particular official figure (let's say about inflation) is unreliable, it would be extremely difficult to change community members' mind using the arguments usually used in statistical circles. Of course, the system also works to underline the validity of figures or sources. Just to underline how this approach is typical of new Internet platforms, the developers of Wikipedia have recently proposed to build a discovery system based on "trusted user feedback from a community of users acting together in an open, transparent, public way". In other words, the proposal is to replace Google discovery algorithms with a system based on the "recommendations" provided by users. This would represent a great challenge, but also a key opportunity, for statistical data providers, who should develop a new communication strategy to convince the whole Internet community to recommend official statistics instead of other sources.

The real question here is: are official data providers ready to engage themselves in this "new world" and therefore to invest significant resources in new forms of communication? For example, if web 2.0 platforms are a marketplace for discussion, and not just a repository of information, should statistical institutions not create discussion sites about the quality of data used in the public domain, including that of their own data? Of course, this could open a "Pandora box" and give space to those who criticise official data, but on the other hand it would allow statistical offices to be perceived as transparent institutions, as well as to express their criticisms on unreliable data

---

discuss ideas, etc. A *digital native* is a person who has grown up with digital technology such as computers, the Internet, mobile phones and MP3. A *wiki* is a medium which can be edited by anyone with access to it, and provides an easy method for linking from one page to another. Wikis are typically collaborative websites, though there are now also single-user offline implementations.

[10] According to Wikipedia, "collective intelligence is a form of intelligence that emerges from collaboration and competition by many individuals" and it can be applied to several fields, such as cognition (market judgments, prediction of future economic and social events, etc.), co-ordination (collective actions, communities interactions, etc.) and co-operation (open source development, etc.). The study of collective intelligence may properly be considered a subfield of sociology, business, computer science and of mass behaviour, a field that studies collective behaviour from the level of quarks to the level of bacterial, plant, animal and human societies.

produced by other sources, as stated by one of the Fundamental Principles of Official Statistics adopted by United Nations (Principle 4: "The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics"). This proactive (and courageous) approach would certainly be coherent with the idea of making the statistical agency a "knowledge builder" for the whole society, putting its unique technical capabilities at the service of the whole society, helping it to discriminate between good and bad information and thus gaining a stronger legitimacy.

5. OECD recent experiences

Over the last two years, the OECD has decided to experiment new tools to make its statistics more accessible and re-usable by users, as well as to test new approaches to communicate statistics and engage people in exploring data and sharing their findings. The following actions have been undertaken:

- In 2006, the OECD Council endorsed a new policy for dissemination of statistics, which involves the re-organisation of statistical products in three broad categories: *OECD Facts and Figures*: a series of simple tables, with commentary, aimed at non-specialists and specialists, to be freely available to all; *OECD Core Data*: up to 1000 ready-made tables, with metadata, drawn from all OECD databases, aimed at students, informed and specialist audiences, to be freely available to all; *OECD Statistics*: a portal giving access to all complete OECD databases, to be available on subscription using the free-at-the-point model[11].
- 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[12].
- In 2007, the OECD made available the data published in its "Factbook" (a selection of more than 200 economic, social and environmental indicators) on Swivel.com, a web 2.0 platform for uploading, exploring, sharing data and disseminating their insights via email, web sites

[11] A key point of this strategy 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.

[12] The OECD is working 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.

and blogs. To manage OECD data, Swivel created a special label "Official Source" to distinguish data uploaded by Organisations like the OECD and by individuals.

- In co-operation with the Gapminder Foundation (www.gapminder.org), the OECD is planning to upload "2008 Factbook" data on Trendalyzer, the software developed by Hans Rösling and his team. The OECD is also planning to create video clips where analysts would present "stories" about countries' performances, policy reforms, etc. based on Factbook data and the use of Trendalyzer and other dynamic visualisation tools.
- As one of the key issues for organisations that produce statistical indicators concerning countries' performances is 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 OECD developed *Dynamic Country Profiles* based on 32 indicators derived from the Factbook. These profiles are represented through four (two concerning economic dimensions, two social dimensions) dynamic "spider charts" (or "radar charts"), where 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, moving averages are used to show how the country situation evolved over time in comparison with the OECD total or other countries.
- In March 2008, the OECD Development Centre launched *Wikigender* (see www.wikigender.org), the first "wiki-based" OECD initiative which aims to facilitate the exchange and improve the knowledge about gender-related issues around the world. A special section is devoted to statistical evidence, where "official" and unofficial data can be easily recognised and evaluated by the audience. In this respect, Wikigender serves as a pilot for the proposed development of a "wiki-progress", in the context of the Global Project on "Measuring the Progress of Societies" (see www.oecd.org/oecdworlforum).

6. Towards a paradigm shift

Some people may argue that all the "signals" mentioned in this paper can be interpreted as being part of a "storm" and not as indicators of a paradigm shift and that there is no need for a radical (and quick) change in the way official statistics are disseminated and communicated[13].

[13] In *The Structure of Scientific Revolution* (1962) the philosopher of science Thomas Kuhn calls "paradigm shifts" the phases of revolutionary change that occur along the development of a scientific discipline when the main

According to several people, we are facing a real paradigm shift[14]. The OECD has recently finalised a report on opportunities and challenges coming from these new technologies and it reached the same conclusion: the web 2.0 is producing a revolution that requires very careful consideration and response[15]. Moreover, it states that "in a world in which people and institutions are 'bombarded' by information every day, each Organisation needs to enhance its role as 'knowledge builder'. To benefit fully from new ICT tools, it needs to develop, with the help of information specialists (including librarians), a true 'knowledge management' approach, by introducing a process to classify information properly, reduce duplication and facilitate dissemination, repurposing and searching. In doing so, it should explore methodologies like crowd sourcing, wikis, etc.".

The claim that statistics is a technique that serves higher public interests and is performed by skilled people in administration compellingly represents the popular image of an "operational discipline" that mostly consists in the production of official data. But now, taken that a discipline's function and target identify its paradigmatic stance, we should recognise that the changes from "information to knowledge" and from "government to society" are relevant enough to demonstrate a "paradigm shift" in official statistics. This revolution comes from the advances in technology, rather than from a new statistical technique: as suggested by Ayres, because of ICT changes, data are becoming a "commodity" and statistical analyses are no longer a kind of methodology whose results are accessible to a small audience, but a key process to produce knowledge for all people.

---

assumptions, values and instruments undergo a radical transformation. In Giovannini (2007), the applicability of this concept to statistics is discussed.

[14] For example, according to Ayres (2007) "we are in a historic moment of horse-versus-locomotive competition, where intuitive and experimental expertise is losing out time and time again to number crunching. In the old days, many decisions were simply based on some mixture of experience and intuition. Experts were ordained because of their decades of individual trial-and-error experience (…) Now something is changing. Business and government professionals are relying more and more on databases to guide their decisions (…) Super Crunchers (i.e. statistical analyses that impact real-world decisions) are not changing the way that decisions are made: they are changing the decisions themselves".

[15] The report states that "the new wave of ICT tools is changing the work of organisations and societies, by enabling individuals to build networks, driven by different values and working styles (horizontal instead of vertical, participatory instead of authoritarian, bottom-up instead of top-down, circular instead of sequential). These new approaches affect the interaction between organisations and their users/stakeholders". It also underlines that "the new ICT tools have a great impact on the rising generations (the so-called 'digital natives'), those who, in 10-20 years, will be recruited by the OECD. The Organisation's use of ICT can affect its capacity to attract young staff. To increase its efficiency, to facilitate the recruitment and retention of the next generation of experts, the OECD needs to review working methods and hiring policies and introduce practices that are more in line with the way the new generations work".

In this context, communication is not an just appendix of the core business focused on data production, but a key function that can determine the success or the failure of an official data provider. Be open to the dialogue with users using the web 2.0 approach is not a choice anymore: it is a must, especially to ensure that new generations will look at official statistics as an authoritative source.

REFERENCES

Ayres, I. (2007) *Super Crunchers*, John Murray Publishers.

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

Carroll, C. D. (2001), *The Epidemiology of Macroeconomic Expectations*, NBER Working Paper No. W8695. Available at SSRN: http://ssrn.com/abstract=294744.

Carroll, C.D. (2002), *Macroeconomic expectations of households and professional forecasters*, Quarterly Journal of Economics, 118 (1), pp. 269-298.

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.