THE WORTH OF DATA: THE TALE OF AN EXPERIENCE FOR PROMOTING AND IMPROVING STATISTICAL LITERACY

Giovanni A. Barbieri and Paola Giacché
Istat (Statistics Italy), Italy
gbarbieri@Istat.edu

Istat, the Italian national statistical institute, in co-operation with professors of statistics, scientific societies and experts in web communication, produced The Worth of Data, *hypertext materials for promoting and improving statistical literacy. We present the experience from two viewpoints: (i) the process for designing and implementing hypertext; (ii) and the ways selected for improving statistical literacy. The first aspect involved the decision to focus on the concept of awareness: not only as to when and how to use statistical data, but also on how to be discerning about sources, their quality and reliability … The second aspect concerned the language and confirmed that to deliver content in plain language, without losing scientific precision, is indeed a difficult task. To achieve good results, it is necessary to make use of the various skills within a good team. Each expert should give up a little turf and contribute knowledge to attain a common outcome worth communicating.*

INTRODUCTION

The Italian cultural tradition is characterized by a cleavage between humanities and sciences and by the primacy of classical studies. On the other hand, the knowledge society asks for a strong integration of competencies and skills and the above dualism is a hindrance to cultural and professional growth.

The hypertext materials *The Worth of Data* were proposed and completed within this framework, in a joint effort of national official statistics (Istat is an independent public research body and has a long tradition of interaction with the academia) and the world of education (universities and scientific societies), as a flexible set of tools for understanding and putting to use statistical concepts and data. The emphasis is on awareness: the reliability and quality of sources, the application of statistical data and concepts to everyday situations, the meaning of data and their relevance.

THE HYPERTEXT

The Worth of Data is a set of hypertext materials – accessible from Istat's website (http://www.istat.it/servizi/studenti/valoredati/) – prepared to help the general public and non-expert readers to improve their use of statistics and directed at anyone interested in using numbers to describe real life situations and make decisions. The project was undertaken by Istat (Statistics Italy) in co-operation with the Italian Statistical Society (SIS), the Inter-University Centre for Teaching Statistics (CIRDIS), the Department for Geo-economical, Linguistical, Statistical and Historical Studies of the University of Rome *La Sapienza*, Odysseus Communication and S3Opus (two private partners). It benefited from a grant by the Ministry of Education, University and Research, in the framework of a programme for the advancement of scientific literacy.

The inspiration came from similar products prepared by international bodies and national statistical offices in other countries: in particular Canada (*Statistics: Power from Data!*) and Australia (*Statistics: A Powerful Edge!*). *The Worth of Data* is not a course in statistics (each "leaf" is self-contained and can be understood without reading the previous ones), but may be used systematically. In fact, the hypertext is primarily addressed to anyone interested in knowing more about official statistics and its uses, including university students in the humanities or "soft" sciences (where statistics is not in the ordinary curriculum), to pupils in secondary schools and, of course, to their teachers. The modules and tools forming the hypertext lend themselves both to self-paced unassisted distance learning and to traditional learning settings (lesson in front of a class, workshop, seminar).

The hypertext is organized in three sections, which outline possible paths in the exploration of statistics. The main contents are in the central column and are organised in four

chapters (but keep in mind that the leaves are largely self-contained): using statistics to make decisions and evaluate policies (statistical data and decision making; statistics and equal opportunities; official statistics in Italy and in the world); the fundamentals of statistics: collection, production, dissemination (censuses; sample surveys; administrative sources; how data are collected, processed and disseminated); how to rely on the quality of statistical data (principles and norms governing the activity of national statistical offices and international bodies; shared methodologies, classifications, definitions and standards; protection of confidentiality); and how to use statistical data (statistical measures; how to prepare tables and charts; how to write a paper based on statistical data). The side-by-side columns open alternative paths. The left one links to statistical information on a few main statistical areas (prices, labour statistics, short-term indicators …): each area is presented in a clear and concise way, before giving access to the actual data. The right column gives the interested user the opportunity to link to other materials on statistical numeracy available on the web.

OUR MOTIVES

In undertaking the project, we had three main objectives in mind: to contribute to improving and promoting statistical literacy; to increase the trust of citizens, respondents and users in official public statistics; and to promote the critical use of statistical information in everyday life.

1. *Statistical literacy*. As stated in the introduction to this paper, Italy has a strong tradition in the humanities, while quantitative and scientific studies have a lesser status. This prejudice is deeply rooted in the Italian culture of the 19th and 20th Centuries, even if Galilei's *Dialogo sui massimi sistemi* is a prominent example of scientific popularization (written in Italian and not in Latin, it exploits the dialogic form to present different points of views). Statistical literacy suffers from this state of things. At school, the introduction of statistics in mainstream curricula is recent and the emancipation from maths is not complete. Citizens at large have no familiarity with the language and the concepts of statistics, with the result that often they are not able to understand precisely the meaning of the figures measuring economic and social phenomena, even if the decisions of people, business and administrations are growingly based on statistics. Freedom of access to statistical information is a way to attenuate information asymmetries (Stiglitz, 1981) and contribute to the progress of a democratic knowledge society, but is ineffective if citizens are not literate, i.e., if they lack the tools to read, understand and analyse statistics; these tools are acquired in school and in lifelong learning, with an interaction between different subject matters and between theory and practice. This set of tools is necessary to professional users (to statisticians, but also to biologists, physicians, sociologists, economists …), and the gap is being bridged at the curricular level. But we think that this is not enough: it needs to be a shared asset, to accustom people to rational and informed decision making.

   The hypertext makes an effort to develop critical awareness in the use of statistical information, that is the ability to discriminate between reliable and quality-certified statistical information (like official statistics tries to produce) and the results of polls (where the methods used in the production process are seldom disclosed).
2. *Increasing the trust in statistics*. Generally speaking, *official statistics* are those produced by national statistical institutions and related networks. In Italy (but also in other national contexts) this definition shows a few shortcomings: the *subjective* facet prevails on the *objective* one and there is a partial overlap between *official* and *public*. From this point of view, official and public statistics are equated when their production is organized by a public body (i.e., the National statistical office, central and local administrations …). In Italy, it could be said that all statistics produced within the framework of SISTAN (the National statistical system) and established in the three-year National statistical plan share the attribute of being official (but not public in the subjective sense, because private bodies may be admitted in the SISTAN network). More substantially, not all SISTAN offices have an effective production capability.

   The second aspect is the *objective* one: official statistics deals with issues sharing a public character *per se* (health, education, environment …) or relevant from the viewpoint of public

information (economic output, energy, credit …) (Stiglitz, 1999). Here, too, the mention of the National statistical plan is important: statistical surveys included in the plan are the result of a complex programming exercise and reflect a national priority (even if it includes statistics required by international agreements and European Union regulations), respect established quality rules and, in most cases, are compulsory for the respondents. In a democracy, establishing (*ex ante*) which statistical information is to be collected is an assurance of its relevance (*ex post*), because it identifies the issues and subjects that the majority of people want to be informed about.

A central feature of official statistics is indeed its quality and, as a consequence, the underlying organization of statistical production. Official statistics is characterized by a stable organization, by the employ of an adequate volume of human, capital and financial resources, by an ethical and deontological code, by the compliance with rules established and acknowledged by an international community of institutions and peers and embracing concepts, definitions, classifications, methods and so on.

Quality and relevance are the pillars of trust in public statistics. On one hand, the quality of the production process and of the products and services disseminated to the public – where official statistics still has an edge, even if its monopoly on statistical information is over – play an important role in establishing trust (Economic Secretary to the Treasury, 1998). On the other hand, the actual use of data and the awareness of users contribute in augmenting it: when information is plentiful, the usual inverse relationship between scarcity and worth is not valid. On the contrary, familiarity and reputation come into play. Users value most the information coming from a source that, in the past, has proved of being trustworthy, dependable, able to provide relevant and good-quality information based on shared standards.

3. *Critical use of statistical information in everyday life*. In Italy, especially in public life, saying tends to prevail over doing, enunciation over results. These habits do not help to spread a statistical attitude, oriented towards finding data, collecting them whenever not already available, using them in decision making, building them while planning and implementing (production or administrative) processes, exploiting them to introduce adjustments, and drawing on them for reporting and accountability. Statistics are often perceived as an annoyance: they entail transparency in decision making, they force to compare wishes with constraints, they unveil announcements and promises without a follow-up, they allow the evaluation and control of policies and actions. Dandekar (1977), as a president of the *National sample survey organization* in India, noticed already in the seventies that when policy makers find that the statistical data disagree with their opinions, they tend not to trust the evidence or, if more sophisticated, asks for additional data and details.

   For these reasons, we see statistical numeracy as a tool for democracy, as a skill that should be in the cultural baggage of every good citizen. If literacy is the capability of expressing oneself and understanding, so as to be able to relate with other people and eventually administrations, statistical numeracy is the ability to understand, appreciate and use simple symbolical expressions (numbers and charts). We have in mind and present in the hypertext many real-life situations: understanding the meaning of a table or chart in a newspaper, evaluating the message of an election poster presenting statistical data, finding and using the relevant statistical data in making a decision, appreciating the merit of an insurance proposal, being aware of the benefits and possible damages of a drug, evaluating the risks of a surgery on the basis of the relevant literature and cases, and grasping the significance and the implication of the Maastricht indicators of the European Union.

UNDERSTANDING STATISTICS

Understanding and using statistical information is not an easy task. Facts and figures reported by statistics do not sound familiar and numeracy itself is not widespread. We assumed the task of accustoming people to reason about facts and figures in the same way they think and reason about familiar issues in their everyday life.

Scientific information in general (and statistical information in particular) hit the audience without a familiar pattern, so that the mind is not able to build a scheme using already known information and to represent the world through social and cultural metaphors (Lakoff and

Johnson, 1980). This process is necessary for the transition from data (the quantitative representation of a phenomenon captured in the moment it happens or is produced), to information (the comprehension and interpretation of data, stemming from the interaction of social actors and the attribution of meaning on the receiving side), to knowledge (the attribution of value to the information, depending on the perspective of satisfying specific users' needs and using the information in a decision making process).

The United Kingdom has a long history (dating from 1799) of promoting the awareness of science and technology. In 1985, the Royal Society set up a working group to look at the nature and extent of public understanding of science (PUS). One of the outcomes of the resulting *Bodmer Report* was the establishment of the Committee on the Public Understanding of Science (COPUS). In 1993, a White Paper committed the Government to support a campaign aimed at raising public awareness of the contributions of science, engineering and technology.

What the British debate taught us is that neither it is enough to inform people about science and its advancements, nor it is correct to attribute the responsibility of the understanding of science to the public, because the scientific and academic institutions have the responsibility of finding the right way to communicate, "translating" from scientific technical jargon into plain language and offering tools for discriminating relevant information (from the user's point of view). Even this effort is not sufficient: science tends to be too specialized and meta-information is not always shared. This leads to the conclusion that, to be effective, popularization involves a dialogue, a continuing relationship between scientists and society. The dialogue needs to be informed, structured and inclusive.

> "Public understanding of science" means the understanding of scientific matters by non-experts. This cannot of course mean a comprehensive knowledge of all branches of science. It may however include understanding of the nature of scientific methods, including the testing of hypotheses by experiment. It may also include awareness of current scientific advances and their implications. Public understanding of science has become a shorthand term for all forms of outreach by the scientific community, or by others on their behalf (e.g., science writers, museums, event organisers), to the public at large, aimed at improving that understanding." (House of Lords, 2000)

It has been argued that the phrase "Public understanding of science" implies a condescending assumption that any difficulties in the relationship between science and society are due entirely to ignorance and misunderstanding on the part of the public; and that, with enough public-understanding activity, the public can be brought to greater knowledge. This approach is inadequate: science cannot ignore the evidence of a decline in trust, and rebuilding trust requires improved communication.

It is clear that the conclusion of the debate on "public understanding of science" may be applied without further discussion to "public understanding of statistics."

THE ROLE OF METADATA

Starting from these grounds, we think that the use of metadata is a device for improving statistical literacy. Metadata are essentially "data about data": the Wikipedia entry goes on with the example of a library catalogue card, which contains data about the nature and location of a book. Metadata are at the core of statistical information because they are indeed essential to understanding and using data, and the more so, the more naïve the user is. Metadata provide a key to data, a pattern for avoiding the wrong idea that data are mere facts, and they just happen, without a rule or a pattern, and without a production activity (Best, 2004).

Metadata are a map which displays the coordinates and shows the route: following metadata and using them as a guide brings the user to the exploration of different fields. Metadata are the foundation bricks on which all other bricks are placed and the coherence of the building derives from this structure.

In our contacts with students, we experienced that they show a lack of knowledge specifically about the definitions, classifications and methods adopted by statistics, and also a lack of awareness of the reality that statistics are not "facts," but the result of a production process, of the application of human intelligence and specific techniques to data collection (Best,

2004). So we realized that before presenting the data we should provide the net, the grid through which the data get their meaning and their significance.

One of the "tricks" in delivering scientific content is to make familiar what is unknown through examples, analogies and metaphors: these tricks stimulate intuition, provide an image, an idea closer to everyday life experiences, and a bridge that allows the inclusion of new elements in our mental landscape.

> If you look at the history of philosophy, you see that all the great and influential stuff has been technically full of holes but utterly memorable and vivid. They are what I call "intuition pumps" – lovely thought experiments. Like Plato's cave, and Descartes's evil demon, and Hobbes' vision of the state of nature and the social contract, and even Kant's idea of the categorical imperative. I don't know of any philosopher who thinks any one of those is a logically sound argument for anything. But they're wonderful imagination grabbers, jungle gyms for the imagination. They structure the way you think about a problem. These are the real legacy of the history of philosophy. A lot of philosophers have forgotten that, but I like to make intuition pumps. I like to think I'm drifting back to what philosophy used to be […]. I went on to say that intuition pumps are fine if they're used correctly, but they can also be misused. They're not arguments, they're stories. Instead of having a conclusion, they pump an intuition. They get you to say "Aha! Oh, I get it!" (Dennett, interviewed in Brockman, 1995)

We find a lot of examples about the use of metaphors in the works of the authors of popular science: the prime numbers are the Mendeleev's periodic table of the elements or the heart rhythm of mathematics (du Sautoy, 2004).

Let's try to go on with our own metaphor: the metadata are the scaffolding of statistical information. Gauss always claimed that mathematics was like a piece of architecture. An architect never leaves the scaffolding for people to see how the building was constructed. Once he had built his proof, Gauss removed the graphic scaffolding so that no trace of his vision remained. This was not an approach that helped mathematicians to penetrate his mathematics (du Sautoy, 2004). Of course we trust in an opposite philosophy: showing the users the scaffolding (metadata) means to provide people with all the information needed to understand the operations involved in producing the data, and even to repeat the experiment. This is at the core of the scientific method. So, other mathematicians were right in criticizing Gauss: his attitude in hiding or removing the scaffolding prevented them from understanding the structure, the links, the connections, the path, and the perspective of his work.

It's a funny coincidence (or possibly not!) that the metaphor used by Gauss to justify his way of working has been denied by a contemporary architect, Renzo Piano, when he designed the Beaubourg Building in Paris. He shocked people with a traditional vision of architecture (the one shared by Gauss) just by leaving the scaffolding and the technical infrastructure in plain view, even after the building was completed. To show the structure means to show the path, from the blueprint through the construction phase to the building, illustrating each choice and alternative, so that every step provides meaning. The same can be said of the role of metadata in statistical explanations.

The presence of the scaffolding of metadata is an antidote to the crisis of trust in official statistics: if people are aware of the entire production process, criticisms and doubts can be faced and coped with in a constructive way. Actually, the problem is not avoiding nasty remarks, but, on the contrary, to allow and promote a critical reasoning about statistical data, providing all the tools necessary for carrying out their proper evaluation in context. This is a crucial issue for a national statistical office: most people are exposed to statistical information only through the media, so that they get second-hand knowledge. If the national statistical office does not provide (actual and potential) users with the opportunity to get statistical information directly, by themselves, wrong attitudes may be encouraged. With the availability of technologic means, such as the web, there is no excuse for not doing it.

DISCUSSION: THE POWER OF STATISTICAL STORY-TELLING

Stories are the natural way to transmit information: our mind simply works this way. Through stories, we build thought representations, which are important cognitive tools, because

they structure and organize experiences (this is the value of metaphors: again, Lakoff, 1980). Stories are also a way to create memories, as shown by the above examples of "intuition pumps."

The popularization of statistics has a lot to learn from story-telling (UN Economic Commission for Europe is working on a *Guide* on statistical story-telling). To deliver statistics by telling stories means to transform facts and figures, by giving them a face, a gender, an age and so on, and describing a fictional "real person" whom we could meet in our country or in our neighbourhood. Writing in order to capture the attention of the reader (the problem is not that the public is not interested in statistics; it is that statistics is not able to compete in catching the eye of the public in a society overwhelmed by information) is not the same as writing a scientific formalized essay: scientific literature advances through hypotheses and demonstrations, popularization through tales legitimated by their likelihood (Carrada, 2005). To conclude, a few recommendations drawn from our experience with *The Worth of Data*:

1. Scientific popularization is a challenging task: it is necessary to find a balance between plain and correct language, without taking anything for granted and always translating technical terms into familiar words.
2. Use the classical 5 W + H (what, who, when, where, why and how).
3. Make a map, first of all for yourself (telling a story as the shared exploration of a territory).
4. Always provide a basic path, affordable by everybody, simple but not *too* simple (as Einstein loved to say).
5. Give the orientation tools for the journey (definitions, glossary, hyperlinks …).
6. Create the opportunity for a more in-depth exploration, through the suggestion of further readings.
7. Structure your material in the form of an inverted pyramid (difficulties and details should be faced step by step).
8. Inform the readers in advance that they are going to meet something more difficult.
9. If you cannot avoid technical terms, explain them at once and in context. Communicating is sharing the language.
10. Always leave the reader some space for serendipitous self-discovery!

REFERENCES

Best, J. (2004). *More Damned Lies and Statistics: How Numbers Confuse Public Issues*. Berkeley and Los Angeles: University of California Press.

Brockman, J. (1995). *The Third Culture: Beyond the Scientific Revolution*. New York: Simon and Schuster.

Carrada G. (2005). *Comunicare la Scienza. Kit di Sopravvivenza per Ricercatori*. Milano: Sironi.

Dandekar, V. M. (1977). Statistics in national planning. *Science Today*, December.

du Sautoy, M. (2004). *The Music of the Primes. Searching to Solve the Greatest Mystery in Mathematics*. New York: HarperCollins Perennial.

Economic Secretary to the Treasury. (1998). *Statistics. A Matter of Trust. Presented to Parliament by the Economic Secretary to the Treasury by Command of Her Majesty, February 1998*. London: The Stationery Office.

House of Lords. Science and Technology Committee. (2000). *Science and Technology - Third Report*. London: The Stationery Office.

Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago.

Stiglitz, J. E. (1999). Knowledge as a global public good. In I. Kaul, I. Grunberg and M. A. Stern (Eds.), *Global Public Goods. International Cooperation in the 21st Century*. NY: Undp.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3), 393-410.

WEB REFERENCES

Statistics: Power from Data!: http://www.statcan.ca/english/edu/power/toc/contents.htm
or *Les statistiques : le pouvoir des données!* http://www.statcan.ca/francais/edu/power/toc/contents_f.htm

Statistics: A Powerful Edge!:
http://www.abs.gov.au/websitedbs/D3310116.NSF/4a255eef08309e44a255eef00061e57/5b6a2b726a507a1f4a2567ac00141fdd!OpenDocument

PUSET: http://www.dti.gov.uk/ost/ostbusiness/puset/puset.htm