

MY SCALE OR YOUR METER? EVALUATING METHODS OF MEASURING THE INTERNET

Giampiero Giacomello and Lucio Picci[#]

This version: September 2002

Abstract

Measuring the Internet - the size of its infrastructure, how many people use it, and their prevalent uses - is of obvious interest. However, the wealth of available quantitative information regarding the Internet so far has fallen short of satisfying the many needs that it would fulfill.

We set the problem of measuring the Internet into a framework that allows us to derive insights on the peculiar nature of the Internet as a piece of infrastructure. After reviewing the current measures available, while drawing a distinction between the object of measurement, and the types of institutions involved in it, we provide some indications on what data should be trusted more, and how better measures of the Internet could be obtained.

Keywords: Internet, measuring the Internet, infrastructure, public capital.

JEL codes: C81, C82, H54, O31

[#] Giampiero Giacomello: Department of Social and Political Science, European University Institute, via Roccettini, 9, I-50016, S. Domenico di Fiesole (FI), Italy. Email: giampiero.giacomello@iue.it; Lucio Picci: Department of Economics, University of Bologna, Strada Maggiore 45, 40125 Bologna, Italy. Email: lpicci@ei.unibo.it, Web <http://www.spbo.unibo.it/picci>. The authors thank for the comments received Larry Press and other participant at the Internet Society INET 2002 Conference (Washington, DC June 13, 2002), Alessandra Colecchia and Hans Engelbrecht.

*“In the Internet era,..., the need for
a new measure is emerging”
(Franklin Daniel, 2001:116)*

1. Introduction

“What counts on the Net?” asked Larry Press, computer science professor at California State University, opening a special issue of the magazine *iMP* (Press, 2000) dedicated especially to measuring the Internet. The appreciation of how “big” is the Internet has considerably risen in the second half of the 1990s, as the Net became an object of much interest at least in the industrialized world. Yet, despite increasing awareness about the phenomenon, determining the size and other features of the Internet has remained unsophisticated at best.

One example of this state of affairs is provided by the rate of growth of Internet traffic, estimated by much press - at least before the collapse of the "dot.com" stocks in 2000 - to double “every three months”. Despite such claims, “...there have been no hard data to substantiate it” (Odlyzko, 2000). Indeed, in the same paper Odlyzko remarks that the belief that Internet traffic could continue doubling every three months *ad eternum* shows “the lack of simple quantitative reasoning” and, ultimately, a case of “innumeracy”.

Given wildly diverging "data" on the Internet, the public, governments and technologists by now are probably wondering whether the Internet will ever be measured precisely. Clearly, there are serious implications in this state of affairs. Companies, public administrations and governments often base their decisions regarding new technologies precisely on those data whose reliability is shaky¹. These measures are important, to the point that some commentators have proposed to use them as supplements to the more traditional ways of measuring national economies².

Not knowing exactly how big is the Internet is also a serious drawback for any scientific research aiming at measuring the impact of the Internet itself. Such researches often use econometric tools to provide an estimate, for example, of what percentage increase in output follows a one percentage increase in a given input, such as labor, or a piece of infrastructure. With the right data, these studies can be carried out both at the economy wide (or "macro") level, and at the firm (or "micro") level. Without such data, the mere possibility

¹ At times, private investors are inexplicably negligent (to say the least) about the authenticity of “figures” from the Internet. More surprisingly, it seems that business people sometimes let themselves to be easily fooled by the same figures. See for instance Judge (2002).

² For example, Franklin Daniel (2001) proposed an “Electronic National Prowess” index.

of measuring the impacts of the new phenomenon, and, as a consequence, of carrying out sound policies, is left to mere deductive reasoning or, even worse, to impressionistic assessments. A major effort is needed to get better measures of the Internet, if we wish to be able to carry out more informative analysis of the “new economy” – and of the “new society” – that the Internet is shaping, and if we believe that good analyses, based on reliable data, are the basis of sound public policies, of sensible business decisions and, ultimately, of public welfare.

There are at least three reasons to explain why, in an age when information of a quantitative nature is ubiquitous, with the Internet we do get numbers, but often not very meaningful ones.

The first one has to do with one characteristic of the Internet that sets it apart from other types of infrastructure. Economists and economic statisticians tend to distinguish between infrastructure and the use that economic agents make of them. On the one hand we may have roads, measured in kilometers, or in an actualized sum of the money that was spent to build them. On the other, we have traffic on those roads.

With the Internet it is often hard to decide what is the infrastructure and what represents a use of the infrastructure. A Web server represents a use of the Internet, but it can also be a "ramp" to access the infrastructure, for example via a Web-based e-mail facility. Also, the same computer that allows uses of the infrastructure - making a home page available, for example - could also be dedicated to purposes that are indispensable for the working of the infrastructure itself, such as the routing of traffic. We then have an object, the Internet, that only with some strain can be put into the existing framework of infrastructure vs. use of the infrastructure.

The second reason is that, even after discounting for this conceptual problem, there is a sheer difficulty in finding the necessary data, in large part due to the decentralized – and, for the most, private – nature of the Internet. Much has been written about the intrinsic decentralization of the Internet, that derives in part from strategic consideration within the US military establishment that first funded it³. However, such a decentralized structure is not incompatible, in principle, with the presence of a single national provider of at least the backbone connections⁴. A single national provider would be in the best position to collect and make public data about the network and its use for one obvious reason – it would control the

³ For a brief history of the Internet, see Hafner and Lyon (1996) and Leiner et al. (2000).

⁴ This was the situation in the United States until 1995, when the Internet was privatized, in concomitance with the National Science Foundation stopping its funding.

whole network, and it could define a unique standard of measurement – and for a less obvious one: once there are more providers in a competitive market, they may have reasons to keep their data private, or possibly even to misrepresent them.

Last, there is what we could define a cultural element representing an obstacle to obtaining good measures of the Internet, having to do with the professional cultures of the people involved. In fact, the professionals who would be in the best position to make many measurements are the engineers who run the infrastructure, who, unfortunately, generally do not have a firm grasp of the economic and statistical issues involved. On the other hand, the people who do have that kind of expertise, for example the statisticians at the national statistical offices, are far away from where many useful measures could be taken, and often do not have a clear appreciation of the phenomenon, nor an understanding of the technical issues that are relevant for coming up with meaningful definitions of what should be measured, and how.

The whole issue of developing good measures of the Internet, however, is best seen from a more detached point of view: A new technological revolution is happening, and an impressive technology adoption process is at work. Economic historians underline that understanding and adopting a new technology is always a long and painful ordeal. They argue that, for example, the second Industrial Revolution of electricity and chemistry actually occurred decades after the relevant inventions took place (David, 1990). In the fields of numbers, state statistics was invented in its modern form in Europe already in the XVIII century, in the age of enlightened absolutism, but we had to wait until the 1930s to see the first systematic data on national accounts (Stigler, 1999).

This paper intends to contribute to this new effort of data construction. Hence it aims to fulfill a trifold goal: (a) to put forward and investigate the important methodological distinction between the Internet as “infrastructure” and the “use” of that infrastructure - or, in different words, between the capital stock proper, and the flow of services that it allows; (b) to provide a “snapshot” of the current status of research in this field; and, as conclusion, (c) to propose some policy recommendations to help address the problem of measuring the Internet. The analytic framework we adopted is broader compared to similar studies (see for instance Engelbrecht, 2001) and allows a clearer understanding of the issues involved. We begin in the next section with examining the distinction between “infrastructure” and its “use.

After that, we provide a review of the main types of data available, first by outlining the different objects of measure, and then by investigating the “who” collects the observations, i.e., national statistical offices, international organizations, etc. We then provide

a series of policy suggestions on how better measures of the Internet could be obtained. Our main conclusion is that measuring the Internet as infrastructure and as “use” of the infrastructure require different criteria and different tools. Both undertakings would be better done by internationally recognized groups, and scientists working on the two project should exchange data and information on the methodology used. Ideally, such groups would include representatives from (a) international organizations (e.g. the UN, ITU, OECD), (b) national statistical offices and (c) non-governmental organizations such as the "Internet governance bodies". The latter point is often missed in other proposals similar to this, but we consider it indispensable to ensure the highest quality of standards for measuring the Net.

2. Infrastructure data and their use

2.1 Infrastructure vs. the use of infrastructure

Economists and economic statisticians distinguish between infrastructure (often also called “public capital”, since it is typically publicly provided, unlike its "private" variety), and the use of infrastructure by economic agents. There are good reasons to keep these concepts separate. One is that, for example, roads are public goods, in the economic sense of the term, and are usually built with public funds, whereas cars and trucks are not. Moreover, infrastructure very often are networks and natural monopolies, in the sense, among other things, that it does not make sense to duplicate a national freeway system in order to encourage competition.

Modern information technologies, by making possible such things as congestion pricing tolls on freeways, may in time put a private touch in many traditionally public goods. However, as of today, traditional infrastructure are still fairly close to the off-the-shelf definition of a public good, with all the implications about their provision and pricing.

Adding relevance to the distinction drawn above is the fact that infrastructure are a type of capital, whereas the decision to buy a liter of gas to drive a car on a freeway is not, and does not add anything to the capital stock⁵.

2.2 Measuring the infrastructure

When it comes to actually measuring infrastructure, two main alternatives are available. Sometimes a physical measure of its consistence is adopted, for example the

⁵ If anything, it takes something away, in the form of wear and tear of the freeway. On public

kilometers of roads, of train tracks, or the number of school rooms.

To aggregate these measures, some weighting is often necessary: a six-lane freeway counts more than a one-lane country road. Aggregation is also an obvious problem, whenever a combined measure of different types of infrastructure – say, roads and railroads together – is desired⁶. Besides these physical measures, and particularly among economists and economic statisticians, (public) capital is very often measured using a "permanent inventory" technique. Essentially, this technique involves adding up past investment flows expressed at constant prices, while deducing the value of assets as they reach the end of their service lives. In order to do it, a sufficiently long time series of the investment flow is needed, as well as information on how the prices of investment goods change in time - or, in different words, the investment data must be expressed at constant prices, in order for their sums to be meaningful⁷.

Capital stocks have been computed using the permanent inventory technique for many countries. For example, for Italy there exist estimates at the regional level both for private and for public capital starting from 1970 (Bonaglia and Picci, 2001). Public capital estimates are also available at the provincial level (Picci, 2002), and longer time series of both private and public capital exist at the national aggregate level.

2.3 What to do with the data

Suppose that we had good data distinguishing the Internet infrastructure from its use. In order to obtain some indications on what to do with them, consider how economists often use capital stock data.

Economists use "production functions" as a way to characterize the output of an economy, or of individual firms. Production functions define output as a function of production inputs, such as labor, energy, and the services of capital, both private, and in principle also of the public, infrastructure, variety. As a matter of fact, the availability of permanent inventory data on public capital has allowed many researchers, particularly after a seminal work of Aschauer (1989), to estimate the role of infrastructure in determining output. These studies have used econometric techniques to estimate production (or cost) functions, in order to determine the statistical significance, and the economic relevance, of the infrastructure input.

goods, see, among many others, the relevant chapter in Varian (1999).

⁶ For such an assessment of infrastructure in Europe, see Biehl (1986).

⁷ A benchmark estimate of capital may also be needed. An important distinction is between

Ironically, the big questions that inspired at least the first wave of these studies was provided by the economic growth slow-down that was observed, in the United States and elsewhere, from the 1970's, much before the more recent debate on "new economy" high growth rates. The role played by infrastructure in determining output is still to some extent controversial, but there is a greater consensus pointing to its significance and economic relevance⁸. These studies have focused on the role of traditional infrastructure, for which data are available for a sufficiently long period of time, and so far have not considered the potential role of the Internet.

Just as finding a reason for the growth slowdown ranked high in economists preoccupations in the 1980's and during the early 1990's, later on, as the US economy picked momentum in the second part of the 1990s, the big question become how to explain the observed revival in growth. The concomitant Internet revolution provided a natural suspect. Again, economists reverted to econometric tools, and a large literature developed, that we can conveniently divide into two main strands. In both of them, economists had to resort to investment data, referring to private capital goods such as computers and software, or to the stock that they form, that is, to varieties of private capital.

One strand of the literature involves "growth accounting" exercises, where, under certain conditions, it is possible to split up of the observed economic growth into several factors, to determine, for example, the role played by the labor input, the capital services, etc. What the relation does not explain is usually called the "Solow residual" (Solow, 1957), typically seen as a proxy of technological progress.

One important issue in these studies, when they are carried out at the sectoral level, is whether parts of the economy other than the ICT sector are growing faster thanks to new technologies. If this is the case, then we would conclude that any aggregate effect is not just due to the observed progress in the ICT industry, but (also) to the spillovers of those technologies, and of the related new organizational practices that they enable, to other sectors.

Using national accounts data, many researchers found that the new technologies were indeed responsible for part of the observed economic growth in the US and in other countries. These results, however, did not go unchallenged; the Winter 2000 issue of the *Journal of Economic Perspectives* provides a review of both camps. These studies employed data on the use of Information and Communication Technologies (ICT) collected from several sources,

gross capital stock, and net (of depreciation) capital stock. Details can be found in OECD (2001).

⁸ For an early survey of the literature, see Gramlich (1994). More recent works (on the US economy) include Pereira (2001), Pereira and dos Frutos (1999) and Fernald (1999).

but that typically did not distinguish Internet related investments from more general ICT expenditure⁹.

Besides these macro studies, econometric techniques were also used with data originated from surveys of firms' behavior¹⁰. There are several surveys of this type available internationally, where firms are asked questions related to their performances and choice of inputs. As we will see, these surveys are typically run by the national statistical agencies and also by other public bodies. The data allow for answering the question of whether firms investing more in the ICT technology perform better, and by how much. These data do not always draw a clear distinction between ICT investment in general, and investment that pertains to the Internet proper, even though, as we will see, questions related to Internet usage have started to appear within already existing surveys.

Overall, the vast body of studies that in the last few years have tried to measure the "new-economy", while not unanimously, in most cases found evidence of what Robert Solow was missing in a well-known quote: "We can find computers everywhere but in the productivity statistics" (Solow, 1987).

3. The object of observation

After having considered how economists often use data on investments and new technologies, we now move on to an analysis of several measures of the Internet, from the ones that are more focused on the infrastructure, to the ones related to its use.

3.1 The cables

As it always happens when we are confronted with new objects, many different metaphors have been proposed to describe the Internet. According to a well-known one, the Internet is an "electronic highway"¹¹. Within that metaphor, cables are then the Internet analogue of roads, possibly with a parallel hierarchy: just as there are many options between dirt roads and major freeways, so there are many types and sizes of cables, from copper telephone lines to fiber optics cables.

⁹ More recent references include Basu et al. (2001) and Bailey (2001), both pointing to a significant role of new technologies in the 90's growth revival.

¹⁰ See, among others, Bryinjolfsson and Hitt (2000)

¹¹ The celebrity of the "electronic highway" metaphor rests with its endorsement by the then U.S. Vice President Al Gore when, in 1993, the US administration launched the "Agenda for Action" to implement the National Information Infrastructure. For an account of US policies on the

Focusing on the cables that make up the Internet would apparently be the closest thing to looking at the Internet as a piece of infrastructure. However, and for several reasons, the "electronic highway" metaphor is not very helpful in understanding the infrastructure nature of the Internet. Traditional roads are characterized by a more or less fixed (in time) relationship between the type of road and the traffic it can sustain. This is not the case for "electronic highways". A good example is provided by the telephone lines used for dial-up, and now also xDSL, connections from home. Dial-up connections speed increased by over an order of magnitude in about ten years, as modem technology improved. XDSL technologies then pushed the limit much further, to the point that, in just one decade, the Internet traffic feasible over a telephone line increased by over two orders of magnitude.

These limitations notwithstanding, a knowledge of the quantity and of the types of cables used for Internet traffic would still provide very useful information to at least get a rough idea of the size of the Internet. The role of technological progress would limit the possibilities of comparisons in the time dimension, but, given that the same technologies are today available at least in principle all over the world, it would not preclude cross sectional comparisons, that is, comparisons among different countries, regions, etc, at the same point in time. Also, an assessment of the quantities of cables would somehow provide an upper bound for the overall transmission capacity of the network, to be reached in case the best technological practices were available everywhere and to everyone.

Unfortunately, data on the quantity and quality of Internet cables are not readily available, mainly because many different organizations - mostly private firms - are responsible for the actual investments. In some instances, there may exist some impressionistic information on the kind of Internet connections that are available within a given community - for example, we may know that somewhere there is a "Metropolitan Area Network" with a given data transmission capacity. Where there exist networks dedicated to research and educational purposes, good data about their nature (and sometimes about their use) are often available. However, these dedicated networks only represents a shrinking subset of the whole Internet, and we just do not have any systematic data on the physical consistency of the cables that make up the overall infrastructure.

Part of these cables, moreover, also serve the telephone network, the most significant example being the lines that allow dial-up and xDSL access to an Internet Service Provider¹².

Information Technology, see Samuelson and Varian (2001).

¹² In some countries - most notably in the United States - the cable television infrastructure also provides access to the Internet.

In this case, an available measure of the infrastructure is provided by the number of people who have a telephone service subscription. The rationale for this is very simple: if a person has a telephone subscription, then the cable is also there for a potential Internet dial-up or xDSL access¹³.

In this case we encounter a theme which recurs frequently in measuring the Internet: often we would like to have a measure of the supply of the Internet infrastructure, but we can only proxy it with some measure of the *use* of that infrastructure. This is also the case with Domain Names counts, that we now consider¹⁴.

3.2 Internet Domain Names

IP (Internet Protocol) numbers are the unique 4-part numbers assigned to each and every computer linked to the Internet, such as "137.204.152.151". A "domain" is a mnemonic string for set of IP numbers. For example, the IP numbers starting with: 137.204 correspond to the domain "unibo.it" (the domain of the University of Bologna). Whoever wants to connect a computer to the Internet first has to obtain an IP number, from a naming authority (for example, the Internet Corporation for Assigning Names and Numbers - Ican - <http://www.icann.org>, for a ".com" address), or from someone who has obtained a set of IP numbers to sub-allocate from such an Authority (for example, the University of Bologna, the holder of the domain "unibo.it", who allocated to one of the authors of this paper the IP number 137.204.152.151, after having received from the Italian Naming Authority the whole set of IP numbers whose first six digits are 137.204).

From these basic information, it is clear that domain names are informative with respect to the size of the Internet infrastructure. There is certainly a direct relationship between how many domain names have been assigned in a given geographical area and the size of the infrastructure there: roughly speaking, the more cables there are, and the more complex is the way they connect with each other, the more domain names have to be allocated. However, domain names are also a measure of *demand*. Given the size of the infrastructure, there is a relationship between the number of domain names and the use of the Internet: the more Web sites there are, the more email services, etc, again, the more domain

¹³ This is true only as a first approximation. The quality of the telephone connection influences the maximum speed of transmission in a dial-up connection. Moreover, xDSL may not be available at a given place due to technical reasons or simply because a market for its provision has not (yet) formed.

¹⁴ Also, no permanent inventory techniques estimate of the Internet infrastructure are available, given the lack of Internet infrastructure investment data.

names are allocated.

For domain names counts we then meet again that blurring between measures of the infrastructure, and of the use that people make of the infrastructure that, as we pointed out, is so common with the Internet. However, worse than this, a host of technical reasons dictate that the direct relationship between domain names and size of the infrastructure, or its use, is only approximate.

First, it is not easy to judge the geographic correspondence of a domain name. One reason is that the very popular domain names allocated by ICANN, and available to everyone (the so called "top level domain - TLD's: .com, .net, .org¹⁵) - are not geographic, and the only way to know where the computers using them are is to ask their administrator. Moreover, even the "Country Code TLD" (ccTLD), such as .fr for France, or .it for Italy, are not really "geographic", in the sense that a computer in a ccTLD could physically be anywhere.

Second, not all allocated domain names are actually used. Over the last few years, and particularly so during the ".com" frenzy of 1999 and 2000, in many parts of the world many domains with possible commercial value were taken up by people who were hoping to make a profit by selling them afterwards, or were simply anticipating the same move by other people, possibly for a potential brand value of the domain name. There is no available reliable estimate of the dimension of "cybersquatting", as this phenomenon has been named, even if we know that it was primarily confined to a subset of rich countries (where it may make sense to pay the US Dollar fees required to hold a domain name idle). Moreover, the incentives to cybersquat could depend on several factors related to developments in technology, such as search engines (Gillmore, 2002), or maybe in the future the Semantic Web (Berners-Lee et al., 2001) and is thus hard to characterize and to control for in the analysis.

Third, while to an assigned domain name there used to correspond a single computer, or "host", now to a single host could correspond many domain names, due to "virtual hosting", whereby a single server can effectively act as if it were several hosts (for example, by providing Web sites to several organizations, with their distinct domain names). This, again, is a further element that blurs the identification between number of domain names and the dimension of the infrastructure, or of the part of infrastructure that we may identify with the physical presence of computers.

Notwithstanding these limitations, and pretty much due to the lack of better alternatives, counting domain names is probably the closest available option to measuring the

¹⁵ Other TLD's are available to a set of US institutions only. A new set of TLD's, which includes

size of the Internet infrastructure, and it is also informative with respect to uses of the Net.

3.3 Servers and their use

Computers connected to the Internet and offering services, or servers, are at the same time part of an infrastructure definition of the Internet, since they contribute to its working, and a manifestation of the uses of the infrastructure. Analysis of servers or, more to the point, of the software that they have installed, can be carried out both through surveys to system administrators, and with "bots". A bot is a generic name for a program that, in the case at hand, automatically analyses computer connected to the Internet in order to determine some relevant characteristics. Such automated programs can provide information, for example, on market penetration rates for Web servers - for an example see Table 3 - or of the number of "secure hosts", a family of software that allows for secure economic transaction on the Internet, and that is often taken to signal the presence of economic commerce activities¹⁶.

Such analysis can also provide useful information, more generally, of Web related activities, such as the number of Web services active, and the amount of information offered by Web servers, etc.

3.4 Counting the users

Users of the Internet can either be individuals or organizations. In both cases, measuring their numbers implies to overcome several obstacles. The first one is that, given the decentralized nature of the Internet, there is no unique registry, akin to the telephone white pages, of the people who subscribe to the service. Moreover, at least for individual users, in order to access the Internet there is no need to register to a service, given that an access at least to the Web may be available in the workplace, at a public library, at a friends' house, or through a cellular phone or other hand-held device.

Another fundamental problem has to do with the definition of "using the Internet". The number of people who meet the requirement of "accessing the Web at least once a month", for example, may be significantly greater than the numbers who satisfy the requirement of "checking e-mail at least 5 times a week".

In the case of organizations, there is a very wide range of activities that they may or

.biz, .info and .name among others, have recently been introduced.

¹⁶ Bots also can operate in two logically distinct way. They can "crawl" the net, that is, go from one computer to others linked to it (for example, via links in the pages of a Web server), or they can analyze a sample of servers selected under some criterion.

may not carry out using the Internet. On one end of the spectrum we may have a firm having an email address that is checked from time to time by a willing employee. On the opposite end, consider a big organization that has successfully reorganized itself through a massive use of the Internet and of Internet related technologies. In all cases, no standard definitions for the different uses of the Internet have emerged yet, and this, even regardless of other measurement problems, is an obvious obstacle for the comparisons and the analysis of data.

Turning for the moment to individual users, and barring the possibility of simply counting the number of entries in an imaginary “Internet white pages” set of books, one way to estimate the number of Internet users is, simply, by surveying random samples of people and by making to them the appropriate questions. These surveys, obviously, may ask several questions so that interesting relations between characters – say, use of the Internet and age, gender, or income – can be ascertained.

Another solution consists in using the domain name count that we have considered, multiplying it by a proportionality factor that typically depends on the country, to take into consideration, for example, that in poorer countries a single e-mail account tends to be used by more people, compared to richer countries where individual accesses are prevalent. However, given the very imprecise relation existing between the number of domain names and the various characteristics of the Internet, such estimates, while popular, amount to no more than educated guesses.

Last, information on the number of users can implicitly be obtained from the analysis of log files, as we will discuss¹⁷. For example, a geographic area where Internet users make up many hits in Web server logs, is a place where there should be many users. We will describe the limitations of this technique later on.

Turning now to organizations, estimating how many of them are on the Internet (and doing what) is largely left to surveys. These can either aim at the whole population of interest or, more often, at randomly selected samples. The existing applications of surveys to this purpose are many, involving both the private sector – we have already considered how these data are used to assess the importance of the “new economy” – and the public administrations, where the focus may be, for example, on the assessment of “e-government” practices. Beyond doubts, at least in the industrialized world, the shift is rapidly changing from counting who is on the Internet, to trying to understand what they are doing there, more or less assuming that

¹⁷ Log files are the files where accesses (for example, to a Web site) leave their mark. They provide three basic types of information: what page or service was accessed, when, and by what Internet address (IP number).

all organized entities are already connected to the Net at least to some extent, or will be so shortly.

3.5 What the users do

Activities on the Internet may be characterized in many ways. On the one hand, we may distinguish between uses of different applications, such as the Web. Also, it may be of interest to qualify further these uses, for example to assess whether, within an organization, e-mail is used only for job-related purposes, or also to organize parties after work¹⁸.

A first solution to the problem, again, is provided by surveys. In the present context, these can take several forms. They can be carried out on random samples of the population at large. In fact, the relevant population for a survey of Internet users is the whole set of people who use the Internet, so surveys could be carried out by interviewing a sample of users only. However, the lack of a general directory of users make it hard to select appropriate samples. Also, the problem in defining what a user is, blurs the boundaries of the users' population.

These limitations notwithstanding, surveys of these type have been conducted in a variety of ways, sometimes using email to solicit an answer, and sometimes via Web forms. In most cases, the statistical properties of the sample are dubious to say the least, as is the statistical inference drawn from them. Very often, it is hard to tell the difference between these "polls" and what goes under the jargon name of "straw polls", where self selection invalidates statistical inference.

In some instances, samples of people are selected to willingly install on their computer a dedicated software that effectively tracks relevant behaviors, such as which software is used, at what time of the day, and for how long. The software may also record what sites have been visited, for how long, which plug-ins or other applications have been used, etc. Such technique, in principle, allows for obtaining a data panel, that is, a cross section of individual observations repeated in time.

Again, and also in this respect, a general problem with surveys is that the definitions of the phenomenon that they employ are *ad hoc*, since no common set of definitions has emerged so far. Another problem is that, as we will see, they are often carried out by private institutions who are wary about disclosing the methodological information that would allow

¹⁸ We only consider legal activities. For criminal (e.g. "cybercrime" or "cyberterrorism") or illegal (such as on-line gambling in many countries), or, in any case, hidden (such as pornography) occurrences, the problems of assessing and analysis are greater, and even more controversial. In this respect, see, for example, Robert Lemos, (2002).

an outside observer to judge their scientific rigor¹⁹. In a sense, private companies running these surveys for profit have an incentive to disclose as little information as possible about their methodology and on the characteristics of the data collected, in part to maximize the profitability of the survey itself.

Another way to inquire into people's use of the Internet is by looking at Web log files, that, as we have seen, can also be used to assess the number of users from a given geographical area. In this case the attention is focused on the Web, but given that the Web has increasingly become the general interface to the Internet, arguably this limitation is not necessarily very severe.

While informative, log files analyses have severe limitations. First, they are collected at each site, so that, in order to obtain them on a wider basis, agreements have to be reached with the (usually private) entities who own the sites. Moreover, a log file analysis cannot yield any particulars of the specific users, nor of their preferences or identities. It cannot even determine if, during a log session, one or more users accessed the computer. Last, Web pages are often seen without causing a "hit" on the servers' log file because they had been previously "cached" in "proxy servers"²⁰.

One last element of trouble in analyzing Web logs is the presence of "bots" themselves. Such programs, now ubiquitous on the Internet, leave a mark on the log files, and tend to complicate their analysis since they are not always clearly recognizable as such²¹.

4. The data available.

We have seen what types of measures and of techniques are available to measure the Internet, together with their qualities and shortcomings. We now describe the broad categories of institutions who are involved measuring various characteristics of the Internet, also to consider how their nature intersects with the scope and the quality of the data that they collect. We begin with an assessment of the role of international organizations.

¹⁹ The mere fact that most times only point estimates are published, with no reference about intervals of confidence, indicates that such rigor is generally lacking.

²⁰ These are computers that, in order to economize in bandwidth, and particularly in expensive trans-oceanic one, save pages locally. Casual observations suggests that the uses of "proxies" is less widespread in the US compared to elsewhere, given that a disproportionately high fraction of contents reside in the US.

²¹ Also, log files are simple text files and they can be very easily tampered with. Their analysis has to be based on a relation of trust with whoever is providing them.

4.1 International Organizations

To different extents, all the main international organizations such as the United Nations, the OECD (the Organization for Economic Co-operation and Development), the International Telecommunication Union (ITU), the World Bank and the European Commission statistical office have devoted resources and attention to measuring the Internet. Indeed, one of the principal activities of international organizations is the provision of data that are *internationally comparable*. However, while each of these organizations can in principle decide to run their own surveys to obtain data, in general they have to rely on the information provided by the national offices or by other organizations²². Thus, if national figures are not dependable; if data are not collected nationally on certain features or activities, or if the data are collected, but according to contrasting definitions of the object of analysis, then comparisons at the international level becomes problematic.

First and foremost, among international organizations, the United Nations Organization (UN) has a considerable expertise in gathering data on different countries to make comparisons possible. The Web page of the UN statistical division offers a comprehensive picture of world statistics ranging from education to HIV, to infant mortality, to the Net.²³ To monitor the fulfillment of the UN Millennium Declaration (adopted by the UN General Assembly in September 2000, and describing the role of the UN in the century ahead) (United Nation, 2000), the UN statistical division has assembled a global database on 48 social and economic indicators,²⁴ including the numbers of Internet users and of personal computers. These data, however, are collected by the UN specialized agency for telecommunications, namely the International Telecommunication Union (ITU).

With 189 member states, ITU is one of the oldest international organizations (it was funded in the mid 19th century) and is now the UN agency dedicated to telecommunications, including the Internet. ITU collects data on Internet users and number of computer users per country²⁵, and carries out country case studies.

The World Bank (WB) is another agency that publishes Internet data (particularly the number of users). However, the figures that the WB uses are drawn from ITU and the Internet Software Consortium (see further).

Among the several international economic organizations that are involved in

²² “The global statistical system is founded, in statistical work, at the national level.” United Nations Statistical Division, “About Global Statistics”, <<http://www.un.org/Depts/unsd/global.htm>>

²³ See <http://www.un.org/depts/unsd/>

²⁴ <http://millenniumindicators.un.org/> .

²⁵ <http://www.itu.int/ITU-D/ict/statistics/> .

measuring the Internet, the Organisation for Economic Cooperation and Development (OECD) over the last years has set up an ambitious plan to measure various aspects of the “new economy”. These range include measures of ICT industries (OECD, 2000) of the types used for the studies on the emergence of the “new economy” that we have considered, assessment of E-commerce and measures of the infrastructure. By "measure of the infrastructure", however, the OECD means those measures, such as the number of domain names and analysis of log files of various type, that we have argued to provide in fact an estimate of a mixture of supply and demand elements (OECD, 1998). Many of the data collected by OECD are published in the so called "Scoreboard"²⁶. Part of these studies are carried out by using data developed by other institutions (such as Netcraft or Telecordia for assessing the "size" of the Internet, see below), and part through surveys conducted by OECD. To this purpose, OECD is developing a "model questionnaire" to measure "ICT use and electronic commerce in enterprises". A summary of the various activities, with indications of future plans, is in OECD (2002).

Eurostat, the European Commission statistical office, also provides information on the Information Society, for the most based on data produced by other organizations, such as ITU²⁷.

4.2 National Statistical Offices

Due to their institutional role, an examination of data-gathering organizations can only begin with the national statistical offices (NSOs). The typical output of NSOs, i.e. statistics, provides the picture of a country’s conditions. So important is their official validation of social phenomena and economic performance that some authors (e.g. Minges, 2000) demand far greater involvement of NSOs in helping to measure the Internet. Most NSOs, however, need deep and serious restructuring of their methods to collect data on the Internet and information flows, as well as to improve their measurement of conventional statistics, which some authors (e.g. Engelbrecht, 2001) consider often “indadequate”. Ultimately, NSOs will have to rethink the measurement of the “building blocks” of national accounts to include the information economy.

The NSOs of some countries, typically the ones where the Internet is more widespread, over the last years have begun to think about measures of the Internet. In some instances, they

²⁶ <http://www.oecd.org/EN/document/0,,EN-document-41-1-no-1-17270-0,FF.html>

have already started collecting the data. In this respect, the US Census Bureau is particularly relevant both because it was among the first NSO's to consider the issue, and because of the quantitative relevance of the United States with respect to the diffusion of the Internet.

The US Census Bureau made early available on-line the data on Internet use in 1996, albeit only on the preferences and opinions of users accessing the Bureau. For more detailed data on users, however, even the Census relied on figures collected by private companies.²⁸ More recently, the U.S. Census has devoted a whole section of its Web site to “measuring the electronic economy” (E-stats)²⁹, which contains information on methods, classification systems, and background papers.

In a series of papers (Atrostic et al., 2000, Mesenbourg, 2000, Mesenbourg, 2001), the US Census laid out its strategy with respect to measuring the Internet or, more to the point, since the focus is explicitly on its economically relevant uses, the "Electronic Economy". The US Bureau of Census first developed and adopted some preliminary concepts of "three key components" of the electronic economy: "Electronic business", defined as "any process that a business organization conducts over computer-mediated network channels"; "Electronic commerce" is "any transaction completed over computer-mediated network channels", and "E-business infrastructure", the "economic infrastructure used to support electronic business process and conduct electronic commerce transaction. It includes the capital (hardware, application software, human capital, and telecommunication networks) used in electronic business and commerce". Note, in particular, the broad infrastructure concept, that includes very diverse types of capital, with human capital among them. Also, note that the concept of "E-business" subsumes phenomena that are often given different names, such as "technology enabled reengineering of organizational processes", "E-government" (and E-governance), etc.

Given these broad definitions, the US Census Bureau aims at answering two distinct sets of questions: measuring "the dimension of the electronic economy", and describing the "impact of the electronic economy on businesses, workers, sectors, regions, and the entire economy" (Altrostic et al, 2000).

This activity is carried out in part by setting up new surveys, and in part by adding appropriate questions to existing Census survey. For example, the data on E-commerce³⁰ are

²⁷ <http://europa.eu.int/comm/eurostat/Public/datashop/print-catalogue/EN?catalogue=Eurostat>. See also the indicators at http://www.europa.eu.int/comm/eurostat/Public/datashop/print-product/EN?catalogue=Eurostat&product=100indic__-EN&type=pdf

²⁸ See for instance the data available on <http://www.census.gov/statab/freq/00s0913.txt>

²⁹ <http://www.census.gov/eos/www/ebusiness614.htm>

³⁰ For the latest available, see <http://www.census.gov/mrts/www/current.html>

obtained as part of the Census Monthly Retail Trade Survey, that exists, in its monthly format, since 1951³¹.

4.3 Research institutions

Many research institutions, mostly academic, have conducted research programs whose aim is to collect data on the Internet, either as an end in itself, or instrumentally in order to estimate the impact of the Internet. Many of these research have involved surveys to users. Given the numbers involved, it is impossible to briefly review this field of research. We limit ourselves to two representative examples summarizing the main interests addressed.

On the one hand, several of these institutions are concerned with an assessment of the changes that the new technologies are causing. For example, the UCLA Center for Communication has recently published the "year two" report of its long term "Survey of the Digital Future"³², a multinational inquiry into how the Internet is changing everyday life.

Other research institutions have begun to specifically focus on estimating the technical performance of the Internet, including (and especially) traffic analysis. In this respect, the Cooperative Association for Internet Data Analysis (CAIDA), based at the University of California San Diego, provides an interesting example of an attempt to "create a collaborative research and analytic environment in which various forms of traffic data can be acquired, analyzed, and (as appropriate) shared".

4.4 Internet bodies

One of the characteristics of the Internet that has often surprised the public, and made many of its users proud, has been the ability of loose communities of technologists to informally organize themselves to "govern" the network. Among such Internet non-profit organizations, the Internet Software Consortium (ISC) is responsible for the main domain names count available. The basis for such a count were laid out in an Internet Engineering Task Force (IETF) "Request For Comment" dated January 1992 (IETF, 1992)³³.

Other organizations collecting data on domain names are the area Regional Internet

³¹ See <http://www.census.gov/mrts/www/noverview.html>.

³² <http://ccp.ucla.edu/pages/NewsTopics.asp?Id=27>

³³ The IETF is one of the most important Internet governance international organizations, responsible for its technical working and evolution. A "Request For Comment" is a typical example of Internet governance. While not being an "Internet standard" (see <http://www.rfc-editor.org/rfcfaq.html>), it may become a *de-facto* one if it obtains sufficient consensus.

Registries (RIRs), non profit organizations whose aim is to provide IP services worldwide³⁴. For example, RIPE provides data on the number of hosts³⁵, supplementing those already made available by the ISC. Table 1 reports yearly observations from 1993 of the ISC total host count, taken from the ISC Web site.

Table 1. ISC Host count

Date	Survey Host Count
Jan 2002	147,344,723
Jan 2001	109,574,429
Jan 2000	72,398,092
Jan 1999	43,230,000
Jan 1998	29,670,000
Jan 1997	16,146,000
Jan 1996	9,472,000
Jan 1995	4,852,000
Jan 1994	2,217,000
Jan 1993	1,313,000

Source: <http://www.isc.org/ds/WWW-200201/index.html>

4.5 Pollsters

Many professional pollsters have conducted surveys on Internet usage. The details of these polls are almost inevitably not made public; it is then not obvious whether the quality of those polls would stand accurate scrutiny. While serious pollsters should be expected to do their work professionally, not all pollster are serious, and pollster's reputation is the main way to judge the numbers that they produce.

As an example of a very long list of pollster that survey Internet usage on a regular basis, we take Gallup, arguably the best known pollster in the US and possibly in the World. Gallup undertakes surveys on the most important issues or events that might be of interests for the American public or private sector. Unfortunately, information is available only to subscribers, thus limiting the access of researchers to the data.³⁶

Traditional polls are mostly carried out by means of computer aided telephone interviews. Recently however several pollster have begun conducting on line polls on many

³⁴ There are three of them: ARIN, catering the Americas and Sub-Saharan Africa; APIC, responsible for Asia and the Pacific, and RIPE, in charge of Europe, part of Africa, and the Middle East.

³⁵ <http://www.ripe.net/ripence/pub-services/stats/hostcount/index.html>

³⁶ A helpful FAQs (Frequently Asked Questions) section summarizing the methodological criteria followed by Gallup for its surveys is on <http://www.gallup.com/help/FAQs/poll1.asp>

topics, Internet usage being one among them. The pros of such practices are lower costs compared to traditional computer aided telephone interview (or "CATI") surveys. The main limitation of such practices is the difficulty to come up with samples that correctly represent the population of interest³⁷.

4.6 Other private organizations

Among other private organizations measuring the Internet, of particular relevance are those firms that develop Internet data mainly as a way to help their customers to better know their (potential) customers to better exploit advertising possibilities. Such analyses are typically carried out either by means of polls on samples of the population or, sometimes, by using other surveying techniques such as focus groups etc. Also, as we have mentioned earlier, some firms employ dedicated software that keeps track of the behavior of a sample of users.

More precisely, marketing and advertising researchers mostly rely on server log file analysis, on bots, and on surveys (Measurecast, 2001), usually done via computer aided telephone interviews or with on-line polls.

These analysis are very often prone to the lack of methodological transparency that we have described. As an example, consider the methodological note published by NUA for its well known "How many on line" survey, which cannot be considered a "methodological note" by any scientific standards³⁸. Table 2 presents a summary of the last data available by NUA, taken from their Web site. Engelbrecht (2001) has detected and exposed similar problems with the "Information Society Index" developed by World Times/IDC.

As an example of a survey conducted by regularly using bots, Netcraft publishes data on several aspects of Internet usage, such as the use of Web servers. Table 3 shows the latest data, taken from the Netcraft Web site. Again, it is hard to judge the reliability of such estimate, whose details are not open to scientific scrutiny. Another respected organization for measuring telecom and Internet traffic is TeleGeography. TeleGeography assesses the worldwide flows of Internet traffic and it is credited to be precise (Engelbrecht, 2001).

³⁷ The National Council on Public Polls (NCPP, an association of polling organizations established in 1969), acknowledging the problem, has published a set of 10 guidelines intended to help journalists to avoid major mistakes and blunder when using figures from on-line polls (<http://www.ncpp.org/internet.htm>).

³⁸ See http://www.nua.com/surveys/how_many_online/methodology.html. It has do be added that the survey is presented by NUA as nothing more than an "educated guess".

However, the staff at TeleGeography admits that their measures cannot distinguish among the *different* types of flows, that is whether the bits are intended for email, Web pages, FTP etc³⁹.

Several not-for profit organizations are also involved in trying to measure different aspects of the Internet. A valuable example is provided by CERT Coordination Center (CEERT/CC), which publishes data on the number of security "incidents" reported to them each year. Other CERT offices are "nationally based" (i.e. they monitor incidents on a "national" computer network), usually with governments, universities or private corporations. There is an ongoing attempt at international coordination in this field through associations such as the Forum of Incident and Response Security Teams (FIRST). However, international comparisons of incidents data are still in their infancy, and international organizations statistical offices do not seem to be concerned with this type of information. This circumstance clearly hampers improvements of measurement techniques in this specific subject.

Table 2. NUA's "How many on line" count: February 2002.

World Total	544.2 million
Africa	4.15 million
Asia/Pacific	157.49 million
Europe	171.35 million
Middle East	4.65million
Canada & USA	181.23 million
Latin America	25.33 million

Source: http://www.nua.com/surveys/how_many_online

Table 3. Market servers for top Web Servers - Active sites.

Developer	March 2002	Percent	April 2002	Percent	Change
Apache	9522954	64.37	10509138	64.38	0.01
Microsoft	3966743	26.81	4431875	27.15	0.34
iPlanet	265826	1.80	278775	1.71	-0.09
Zeus	170023	1.15	182918	1.12	-0.03

Source: <http://www.netcraft.com/survey>

³⁹ Personal communication to the authors, Internet Society INET 2002 Conference, Washington, DC June 13, 2002.

An interesting way of providing information on Internet usage and size, as well as users' habits, could be the diffusion of portals like the one of the Pew Internet & American Life project, funded by the Pew Charitable Trust, and focusing on the United States only. These organizations are well suited for providing useful and independent information about the Internet and the Information Society. The Pew Internet & American Life research staff collects and make available survey data, reports and analysis on the Internet.

5. Conclusions: Where should we go from here?

This paper pursued a trifold goal: (a) to put forward and investigate the important methodological distinction between the Internet as “infrastructure” and the “use” of that infrastructure; (b) to provide a “snapshot” of the current status of research in this field; and, as conclusion, (c) to propose some policy recommendations to help address the problem of measuring the Internet. To this end, we have reviewed the principle types of measures of the Internet and, subsequently, the main actors of the data collection process. While doing that, we have listed a host of reasons why those measures are generally unsatisfactory. The broad framework that we have employed to analyze the current situation, besides helping in highlighting the shortcomings of the data, is also useful to suggest directions for future development. So, the question that we now pose is: Where do we go from here? How can the present situation be improved, and who is in the best position to make a contribution?

The first problem that we have highlighted has to do with the conceptualization of the Internet as a piece of infrastructure. We have argued that, in the case of the Internet, the traditional distinction between infrastructure vs. its use is blurred. As a consequence, the term "infrastructure", when it comes to the Internet, is currently used very loosely. We have seen two significant and authoritative examples: the US Census Bureau includes human capital, among other things, in its definition (US Census, 2001), while OECD by "infrastructure data" seems to mean whichever data comes from an observation of the actual working of the Internet, regardless of what it actually measures (OECD, 1998).

Such conceptual ambiguities are not of help when it comes to define the object of measurement. Among econometricians there is the saying that there can't be "measurement without theory"⁴⁰ A conceptual effort in trying to better understand the nature of the Internet as a piece of infrastructure, that is, a better theory, is a prerequisite for making progresses in

⁴⁰ With reference to Koopmans (1947).

obtaining good measurements of the Internet. The burden of such an effort would obviously fall on the shoulders of the academic community.

The benefits in clarifying our thoughts about the infrastructure nature of the Internet may well spillover to the analysis of more traditional types of infrastructure. Even roads are not what they used to be, that is, an unintelligent mass of concrete and asphalt. New technologies are taking over, making possible congestion pricing, advanced forms of planning of infrastructure use, and futuristic approaches to the management of logistics. Computers and networks are being integrated into buildings, to the point that a word, "domotics", has been coined to analyze and describe such a process. Such use of the new information technologies define new types of infrastructure, whose components, just as it happens with the Internet, could be seen as services, as an expression of demand, and at the same time as an integral part of the infrastructure proper. In other words, the analytical ambiguity that we have here witnessed for the Internet is today spreading to more traditional realms. For this reason, the need to update our conceptual framework is all the more urgent.

In order to obtain better data, a further task to be carried out has to do with the existence of common definitions of the objects of measurement, whose lack today is a powerful impediment to comparisons of data, not only internationally. We believe that two types of organizations could help in this. First, there are the international organizations, such as OECD and ITU. We have briefly described OECD efforts both at measurement, and at the developing of a coordinated framework among member countries' statistical offices. Such efforts should be made more intense and visible⁴¹. These efforts should be firmly rooted into the broader issue of international harmonization of statistical information (e.g. Lynn, 2001).

Another interesting possibility rests with the involvement of the Internet governance institutions. These organizations have so far operated on networking standards (the Internet Engineering Task Force, IETF, or the Internet Architecture Board, IAB), Web standards (the World Wide Web Consortium, W3C), on the managing of domain names (Internet Corporation for Assigned Names and Numbers, ICANN), or on Internet public policies (the Internet Society, ISOC). Several of these organizations have published guidelines on various technical matters. However, guidelines on "how to measure" the Internet have been remarkably lacking. The model for such an action could be a Request For Comments scheme akin to the "RFC" that have accompanied the technical developments of the Internet - including, for instance, the introduction of the TCP/IP protocol - or any suitable formalization

of a similar "Internet governance" approach.

A coordinated effort from this camp would have two advantages. First, it would be the best way to build consensus on the need for measurement within the community of technologists that runs the Internet, and that are in the best position for many types of measurements (traffic, log files, etc.). Second, if carried out with the involvement of economists and economic statisticians, it could help put together the different expertise that are needed to obtain good measures of the Internet, and bridge that cultural gap that, as we have argued, is one of the reasons for the lack of good data today.

The lesson of Internet governance should be learned by those international organizations, such as OECD and ITU, that, as we have argued, should be more active in promoting common standards for the collection of data on the Internet. Too often their working is not very transparent to outside observers, and it fails to build the necessary consensus of the interested parties, including, ultimately, the users.

Such criticisms and suggestions for improvement, however, are best seen in perspective. Nobody foresaw that the Internet would become so important. The many fathers of the Internet envisaged it first as a strictly U.S.-based network, then only as a university-based infrastructure. Nobody foresaw millions of individual users and of companies on the Net. Few, if any, among the first Internet users and developers thought about assessing the long-term consequences of the Internet. We should then not be surprised that it took a while before the need for good measures of the Internet affirmed itself.

One of the reasons for the impressive success of the Internet rests in its technological soundness, which is guaranteed by an interesting governance that puts a premium to professional expertise. Engineers have built and are running the Internet, and their main goal has been, and it is, in guaranteeing the working of the infrastructure, not its measurement *per se*. Now it is also the time for economists and economic statisticians to take a bigger role, and to work together with the engineers to come up with better solutions to the problem of measurement.

Units of measure are standards that individuals in society have agreed to use. The collaboration of all these actors is necessary in order for standards to emerge, to be agreed upon, and to be adopted by all interested parties, so that we can obtain the good measures of the Internet that are currently lacking.

⁴¹ Even on the Web. The OECD Web site, for example, does not represent adequately the work done by that organization in the field.

BIBLIOGRAPHY

- Aschauer, D. (1989), "Is Public Expenditure Productive?", *Journal of Monetary Economics*, vol. 23, pp. 177-200.
- Atrostic, B.K, John Gates and Ron Jarmin (2000), *Measuring the Electronic Economy, Current Status and Next Steps*, US Census Bureau, mimeographed.
- Berners-Lee, Tim, James Hendler and Ora Lassila (2001), *The Semantic Web*, *Scientific American*, May 2001.
- Bonaglia, Federico, e Lucio, Picci (2000), *Il capitale nelle regioni italiane*, WP. n. 374, DSE, Università di Bologna.
- Brynjolfsson,-Erik; Hitt,-Lorin-M. (2000), *Beyond Computation: Information Technology, Organizational Transformation and Business Performance*, : *Journal-of-Economic-Perspectives*; 14(4), Fall 2000, pages 23-48..
- Daniels, Franklin (2001), "The Next Measure of National Machismo", *The World in 2001*, London: *The Economist* pp.116/119
- David, Paul (1990), "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox", *American Economic Review*, May 1990, pp. 355/361.
- Engelbrecht, Hans-Jurgen (2001), "Statistics for the Information Age", *Information Economics and Policy*, Vol. 13, Issue 3, September 2001, pp. 339-349
- Fernald, John (1990), "Roads to prosperity? Assessing the link between public capital and Productivity", in *American Economic Review*, 89(3), June 1999, 619-638.
- Foley, Kathy (2001), "Measure for Measure", Editorial, *Nua Analysis*, November 26, http://www.nua.ie/surveys/analysis/weekly_editorial/archives/issue1no205.html
- Glillmore, Dan (Jan 12, 2002), "Google effect' reduces need for many domains", *Sylicon*

Valley.com, , <http://www.siliconvalley.com/docs/opinion/dgillmor/dg011301.htm>

Glasner Joanna, (2001), "Who's No. 1? Depends Who Counts", *Wired News*, November 1st,
<http://www.wired.com/news/ebiz/0,1272,48041,00.html>

Gramlich E. (1994), "Infrastructure Investment: A Review Essay", *Journal of Economic Literature*, Vol. 32, pp. 1176-1196.

Greenstein Shane, (November 2000), "By the Numbers: Musings about Numerical Challenges in the Information Age", *iMP Magazine*, Washington, DC: Center for Information Strategy and Policy, Science Applications International Corporation,
http://www.cisp.org/imp/november_2000/11_00greenstein-insight.htm

Hafner Katye and Matthew Lyon, (1996), *Where Wizards Stay Up Late. The Origins of the Internet*, New York: Simon and Schuster

IETF (1992), Request For Comments n. 1296,
<http://www.ietf.org/rfc/rfc1296.txt?number=1296>

Judge, Peter (January, 9, 2002) ".Net vote rigging illustrates importance of Web services", in *ZDNetUK*, <<http://news.zdnet.co.uk/story/0,,t269-s2102244,00.html>>

Kalton, G. (1983) *Introduction to Survey Sampling*, QUASS 07-035, Beverly Hills, CA: Sage

Koopmans, Tjalling C. (1947) Measurement without theory, *The Review of Economics and Statistics*, Vol XXIX, pp. 161-172.

Leiner Barry et al. (August 2000), A Brief History of the Internet Version 3.31, Internet Society, <<http://www.isoc.org/internet/history/brief.shtml>>

Lemos, Robert (2002). "Security: What's Going On?", *ZDNet News*, 21 January, 2002, <<http://zdnet.com.com/2100-1105-819713.html>>.

Lynn, Peter (2001), “Developing Quality Standards for Cross-National Survey Research: Five Approaches”, *Working Papers of the Institute for Social and Economic Research*, 2001-21,. University of Essex, <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2001-21.pdf>>

Media Measurement Task Force (September 15, 1997), *Metrics and Methodology*, Interactive Advertisement Bureau (IAB), <<http://www.iab.net/advertise/adsources.html>>

Measurecast, (2001), *An Analysis of Streaming Audience and Measurement Methods*, mimeo, August 9, http://www.measurecast.com/docs/Audience_Measurement_Methods1.pdf

Mesenbourg, Thomas L, (2000), *Measuring the Digital Economy*, U.S.Bureau of Census, mimeographed.

Mesenbourg, Thomas L, (2001), *Measuring Economic Business*, U.S.Bureau of Census, mimeographed.

Minges, Michael (2001), *Internet Around the World*, mimeo, INET 2001 Conference, Sweden, http://www.isoc.org/inet2001/CD_proceedings/G54/Inet2001_100501.htm

Minges, Michael (2000), *Counting the Net: Internet Access Indicators*, mimeo, INET 2000 Conference, Japan, http://www.isoc.org/inet2000/cdproceedings/8e/8e_1.htm

OECD (1998), *Internet Infrastructure Indicators*, <http://www.oecd.org/pdf/M000014000/M00014287.pdf>

OECD (2000), *Measuring the ICT sector*, <http://www.oecd.org/pdf/M00002000/M00002651.pdf>

OECD (2001), *Measuring Capital: A Manual on the Measurement of Capital Stocks, Consumption of Fixed Capital and Capital Services*. Paris: OECD.

OECD (2002), *The Working Part on Indicators for the Information Society (WPIIS): progress report and ideas for future work*.

Pereira, Alfredo (2001), "Is all Public Capital created Equal", *Review of Economics and Statistics*, 82(3), August 2000, 513-518.

Pereira, Alfredo, and Flores-de-Frutos, Rafael (1999), "Public Capital and Private Sector Performance", *Journal of Urban Economics*, 46(2), September 2000, 300-322.

Picci, Lucio (2002), "Le infrastrutture in Italia. Le differenze territoriali e l'efficienza della spesa", in: *L'Italia nella competizione globale - Regole per il mercato*, a cura di Mario Baldassarri, Giampaolo Galli e Gustavo Piga, Edizioni il Sole 24 Ore, Milano.

Press, Larry (November 2000), "What Counts on the Net? Frameworks and Issues in Measuring the Internet", *iMP Magazine*, Washington, DC: Center for Information Strategy and Policy, Science Applications International Corporation

Odlyzko, Andrew (November 2000), "Internet Growth: Myth and Reality, Use and Abuse", *iMP Magazine*, Washington, DC: Center for Information Strategy and Policy, Science Applications International Corporation,
http://www.cisp.org/imp/november_2000/odlyzko/11_00odlyzko.htm

Samuelson, Pamela, and Hal Varian (2001), *The "New Economy" and Information Technology Policy*, mimeo

Stigler, Stephen M. (1999), *Statistics on the Table. The History of Statistical Concepts and Methods*, Cambridge, MA: Harvard University Press

United Nations (2000), *The Millennium Declaration*, <http://www.un.org/millennium/sg/report/>