International Journal of Communication 6 (2012), 956–979 1932–8036/20120956

Information Capacity
IJoC

How to Measure the World's Technological Capacity to Communicate, Store, and Compute Information Part I: Results and Scope[1]

MARTIN HILBERT[2]
University of Southern California
United Nations ECLAC

PRISCILA LÓPEZ[3]
University of Catalonia

This is Part I of a two-part article that reviews methodological and statistical challenges involved in the estimation of humanity's technological capacity to communicate, store, and compute information. It is written from the perspective of the results of our recent inventory of 60 technological categories between 1986 and 2007 (measured in bits and MIPS [million-instructions-per-second]). In Part I, we summarize the results of our inventory, and explore a series of basic choices that must be made in the course of measuring information and communication capacities. The most basic underlying assumptions behind our estimates include—among others—decisions about what is

[1] The authors would like to thank Roger Bohn, Andrew Odlyzko and István Dienes for their detailed comments, and United Nations ECLAC for giving us the opportunity to set the groundwork for many of the ideas developed in this article. We would also like to thank Tom Coughlin from Coughlin Associates, John McCallum, Don Franz from Photofinishing News, Joerg Woerner from Datamath, Manuel Castells and Len Adleman from USC, our research contributors Miguel Gonzalez and Cristian Vasquez, and the very helpful statisticians from ITU (International Telecommunication Union), and UPU (Universal Postal Union).
[2] Provost Fellow, USC Annenberg School of Communcation & Journalism; Economics Affairs Officer, UN Economics Commission for Latin America and the Caribbean.
[3] Research Scholar, Information and Communication Sciences Department.

Martin Hilbert: mhilbert@usc.edu
Priscila López: priscila.lopez@gmail.com
Date submitted: 2011–06–11

counted as (1) communication, (2) storage, and (3) computation; if technological capacities or consumption of information is measured; and if unique information is distinguished from duplicate information. We compare our methodological choices with different approaches taken in similar studies. The article shows how the particular question on the researcher's mind, as well as the availability of source data has and will influence most of the methodological choices in different exercises.

Lord Kelvin, who famously revolutionized the scientific enterprise with the introduction of a new indicator, is accredited with the quip "To measure is to know." But measuring something for the first time is not straightforward. In this article we discuss the methodological and statistical challenges involved when measuring the amount of information processed by the world's information and communication technologies (ICT). We start with a review of the results of our account of the world's technological capacity to communicate, store, and compute information (see Hilbert & López, 2011). This helps set the stage and give a sense of the orders of magnitude involved, but we do not fall short in pointing out that these results are not without controversy. There are several ways one can measure the world's capacity to handle information. In the second part of Part I we review some of the basic choices that have to be made when quantifying "the world's information explosion" (Bialik, 2011). We naturally do this from the perspective of our own inventory, but also discuss the trade-offs involved in our choices, and as often as possible refer to the benefits of the choices made by researchers of other studies of a similar nature. Part II of this article analyzes the choice of the central units of measurement of information (which for our exercise are "optimally compressed bits"), and close with an outlook into this ongoing challenge.

Results of Our Inventory

We published the first results of our inventory of the world's technological capacity to store, communicate, and compute information in Hilbert and López (2011). We take a closer look at this total from various perspectives in a series of follow-up studies.[4] This inventory has received an unanticipated amount of media attention from the general public (e.g., Bär, 2011; Bialik, 2011; de Luna, 2011; Greenemeier, 2011; Lebwohl, 2011; Schrader, 2011; Stewart, 2011; Technica, 2011; Vastag, 2011).

Our estimations are informed by more than 1,100 sources (listed in the Appendix: 120 sources for the case of content compression, 360 sources for the case of storage, 380 sourses for the case of communication, and 300 sources for the case of computation). Among the databases that we rely on, it is worthy to acknowledge and commend the important and valuable work done by ITU (2010), which includes indicators on television, radio, Internet, and phone subscriptions; Porter (2005) for an invaluable historical inventory of the world's hard disk drive storage, as well as Coughlin (2007), IFPI (2005), and JRIA (2007) for an account on optical storage and other storage devices, Faostat (2010) for an inventory on the world's paper production, UPU (2007) for an account of postal statistics, TOP500 (2009) for their tracking of supercomputers, as well as Longbottom (2006) and McCallum (2002) for their work on personal and professional computer performance. The Appendix mentioned in this article is an updated version of the Appendix published in López and Hilbert (2011) and is available at

[4] See www.martinhilbert.net/WorldInfoCapacity.html to access these studies (e.g., Hilbert, 2011a).

http://www.martinhilbert.net/WorldInfoCapacity.html. This also slightly changes some of the numbers reported for 2007 (in comparison with Hilbert and López, 2011).

The World's Technological Capacity to Store and Communicate Information

We start by measuring two basic information operations: the transportation of information through (1) time (storage) and (2) space (communication). We measure the first in the installed capacity of optimally compressed bits for a given year, and the second in the number of optimally compressed bits that have been effectively communicated in a given year (more on the particularities of this specific indicator in Part II of this two-part article). This aggregates distinct forms of information (digital or analog; in text, images, audio, video, etc.) into a single variable for both cases.

The world's technological capacity to effectively communicate information has grown from 432 exabytes in 1986 (4.3210^20 bytes) to 1.15 zettabytes in 2007 (1.1510^21 bytes). The amount of information stored grew from 2.6 exabytes in 1986 to 309 exabytes by 2007. On the one hand, these are impressive amounts of information. If we wanted to use grains of sand to represent one bit of information that was communicated in 2007, we would require 2,000 times the number of grains of sand on the world's beaches (compare McAllister, 1997). On the other hand, it is still comparatively small. It is estimated that the amount of information stored in the DNA of the 60 trillion cells of one single human adult is at about the same order of magnitude (10^22 bytes)[5] (see also Hilbert, 2011b).

Figure 1 shows that the majority of the world's technologically mediated information has been broadcast during the past two decades (measured at the left vertical axis in the 10^15 range, compared to storage, measured at the right vertical axis in the 10^14 range). However, the undisputed dominance of broadcasting is slowly but surely being challenged. The lion's share of the global flow of information is provided by the bits that are relentlessly transmitted through the channels of television technology (between 93–95% of all technologically mediated communication). The world's broadcast capacity has grown with a compound annual growth rate (CAGR) of some 7% during the two decades covered. Telecommunication has grown much faster, at 28% per year, but communicates a much smaller amount of information (less than 3% of the total amount of bits communicated in 2007). The growth rate of telecommunications is similar to the expansion of the world's technological storage capacity, which has grown at a compound annual growth rate of some 25% per year.

In all cases, it becomes clear that the process of digitization has grown decisively, especially since the year 2000. While only 25% of the information flow of broadcast networks had been digitized in 2007, the digitization of storage and telecom is almost complete (94% and 99.9% digital in 2007, respectively) (see Hilbert & López, 2011).

[5] There are roughly 3 billion DNA base pairs per human cell and 60 trillion cells per adult human. DNA consists of a quaternary alphabet, which would require 4 bits per base pair, but since base pairs are predetermined (A goes with T, G with C), one can compress this to 2 bits per base pair. Biologists use several compression algorithms to compress this information inside one cell, but these only achieve moderate results (see for example Chen, Li, Ma, and Tromp, 2002).

Figure 1 also clearly shows that every year we communicate much more information than we can possibly store. While the capacity of two-way telecommunication and our storage capacity have grown at similar growth rates, broadcasting has grown more slowly. Since broadcasting represents the vast majority of transmitted bits, the respective ratio of [storage]**/**[broadcasting**+**telecom] is changing in favor of storage. In 1986, we could have stored less than 1% of all the information that was communicated around the world in all our technological devices (including paper, vinyl, tape, and others). By 1993, this share increased to 3%, to 5% by 2000, and to 16% by 2007. The role of storage is often underestimated, as much of the literature on the so-called information and network society focuses almost exclusively on bidirectional communication networks (i.e., Internet and mobile phones) (e.g., Castells, 2009; ITU, 2010; Mansell, 2009). However, as the results of this inventory show, just as important as a key characteristic of the emerging Information Society is the storage of information in ever-increasing digital memory.

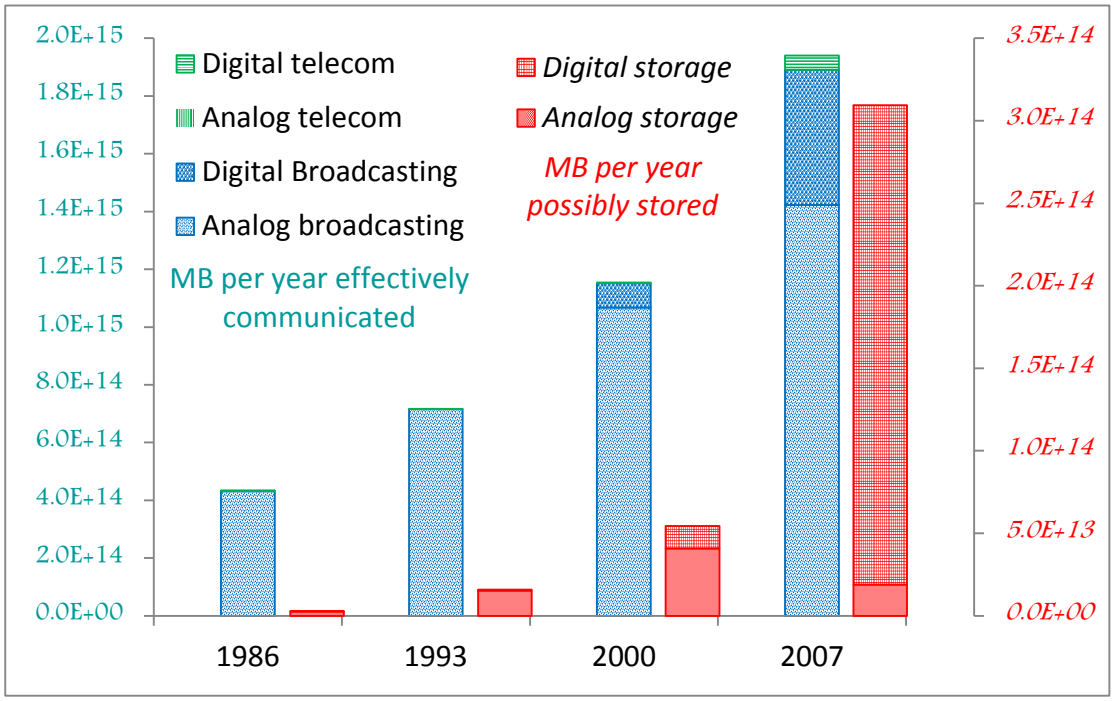

Figure 1. Total sum of technologically mediated information: Installed capacity of storage, and effective capacity of broadcasting and telecommunication, in optimally compressed megabytes (MB) per year for 1986, 1993, 2000, and 2007.

Source. Authors' own elaboration, based on various sources.
See Appendices, also see Hilbert and López, 2011.

Therefore, one of the straightforward conclusions of this exercise is that information storage plays an important and often neglected role in the digital age. While technological communication is often the main focus of research on the information and network society,[6] Figure 1 shows that it is the storage of information in vast technological memories that has made a most impressive catch-up. Some speculate that the future will bring the long-expected scenario where people no longer possess a significant, permanent kind of physical storage device, and all content will be stored "in the network cloud," conveniently accessed through unlimited bandwidth in real-time (Cloud storage, 2010; Etro, 2009; Meyer et al., 2010).[7] For now, however, we have to recognize that during the past two decades humankind has started to hoard information in its personal technological memories. In other words, the digital age increased our capacity to store information (and often "copy-paste" duplication) much more than our capacity to transmit information through broadcast and telecommunication networks.

The World's Technological Capacity to Compute Information

In addition to transmission of information through space and time on a large scale, we also employ technologies to transform it meaningfully through some kind of logical procedure (an algorithm). We measure the hardware capacity of computers in MIPS (millions or mega instructions per second), and distinguish between two kinds of computers: general-purpose computers and application-specific computers. The first kind refers to what we generally understand under the term "computer," such as is used in everyday language, and includes all kinds of humanly guided computers. The user can decide what to do with the computing power and can deliberately employ it for tasks such as word processing a document, watching a movie, or executing a mathematical calculation. This kind includes personal computers (including laptops), servers and mainframe computers, supercomputers, videogame consoles, mobile phones and handheld personal digital assistants, and pocket calculators. The second kind refers to computers that are specialized in a specific task and embedded in some kind of machine, such as an electronic device or household appliance. This includes digital signal processors (DSPs), which relentlessly transform between analog and digital signals (such as in CD players and modems), microcontrollers (MCUs), which can be found in everything from a microwave to a car, and graphic processing units (GPUs), which control the correct display of information on a monitor screen.

While our inventory of general-purpose computers is complete, our sample of application-specific computers is more limited, since this group of computers is large and diverse. In theory, every dice cup and roulette wheel is an application-specific analog random number generator that should be included in a complete inventory of this group. There are also additional, simple logic gates and other semiconductors that we did not include in our inventory of application-specific computers. Our main ambition in this

[6] A quick search at the online bookseller Amazon.com (the world's most popular online retailer in this rubric in 2010), shows 54,000 book results for the search term *telecommunication*, 28,000 for the term *broadcasting*, and only 12,000 book results for the term *information storage* (search without quotation marks) (executed Dec. 30, 2010).

[7] One indication for this scenario is the fact that the world's technological memory is ever more concentrated among devices. Large storage devices capture an ever larger share of the world's total capacity to store information.

exercise is to show that the world's technological capacity to compute information through application-specific computers is larger than the capacity of its general-purpose computers, and that this difference is growing. To show this it is sufficient to focus on three of the most prominent application-specific computers: DSPs, MCUs, and GPUs (see Figure 2).

The combined capacity of both groups of computers has grown from 730 tera-IPS in 1986 (73010^12 instructions per second), over 22.5 peta-IPS in 1993 (22.510^15 instructions per second), to 1.8 exa-IPS in 2000 (1.810^18 IPS), and 196 exa-IPS in 2007 (or roughly 210^20 instructions per second). In comparison, the estimated age of the universe is roughly 410^17 seconds since the big bang. In 1986, general-purpose computers still represented 41% of the world's computational capacity, while this share has continuously shrunk to a mere 3% by 2007. Until the mid-1990s, microcontrollers (MCUs) in electronics and appliances provided the majority of the world's computational power; since then, GPUs took over. These consist of integrated or discrete graphic cards, which are specialized microprocessors that accelerate graphics rendered from the central processing unit (CPU) of a general-purpose computer. Their highly parallel structure makes them more effective than general-purpose CPUs for a range of complex algorithms.

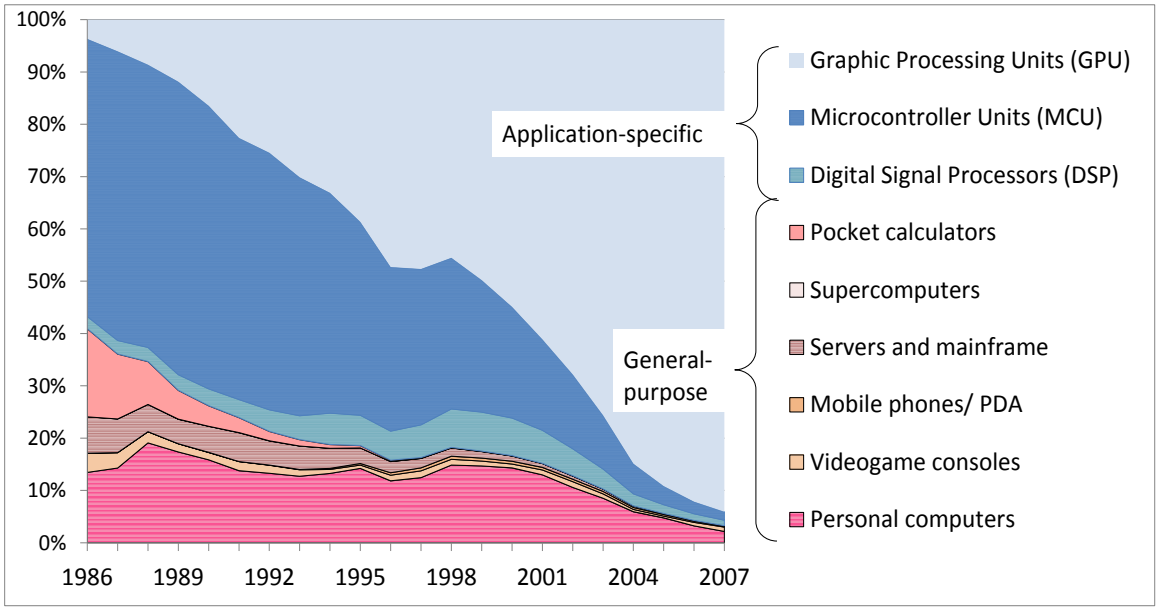

Figure 2. Total sum of the installed technologically mediated computational capacity: Application-specific and general-purpose computers, share of million instructions per second (MIPS) for 1986–2007.

Source: Authors' own elaboration, based on various sources.
See Appendices, see also Hilbert and López, 2011.

One of the straightforward conclusions is that computers have increasingly more computational capacity to process information among themselves (through embedded application-specific computers) than we human users can compute with our machines (on general-purpose computers). Another conclusion is that our capacity to compute information has grown much faster than our capacity to communicate or store information. Between 1986–2007, our technological capacity to effectively broadcast information has grown at a rate roughly equivalent to the world's economic growth (global GDP has grown with 6% per year during the same period). The world's installed capacity to store information has grown four times faster than the economy, and the world's effective capacity to telecommunicate information has grown almost five times faster. The respective compound annual growth rates between 1986–2007 were 61% for general-purpose computations and 86% for application-specific computations, which is 10 and 14 times faster than global GDP during that period, respectively.

Just One Way of Looking At It

While the presented numbers give an interesting overview of the orders of magnitude involved, it is important to remember that they are only one way to quantify the amount of information processed by the technologies of a society. In line with the other articles in this Special Section, we compare and contrast our approach with the methodologies applied in previous work (see, i.e., Bohn & Short, 2009; Bounie, 2003; Gantz et al., 2008; Ito, 1981; Lesk, 1997; Lyman, Varian, Dunn, Strygin, & Swearingen, 2000; Lyman et al., 2003; Neuman, Park, & Panek (see this Special Section); Pool, 1983; Short, Bohn, & Baru, 2011). We underline the specific assumptions made in our inventory, consider alternative ways that could have been taken to present them, and discuss some tradeoffs between the different forms of presentation.

Scope of Our Exercise: Basic Choices of What to Measure

We decided to estimate the world's technological capacity to communicate (in optimally compressed bits per second), store (in optimally compressed bits), and compute information (in MIPS). This implies that we:

(a) only focus on technologically mediated information;
(b) distinguish among communication, storage, and computation;
(c) focus on the technological capacity (not consumption or uniquely created information);
(d) normalize bits and bits per second on compression rates and, therefore, convert hardware capacity into the respective amount of information.

The first choice (a) is justifiable because it can roughly be assumed that (in quantitative terms) the biological information processing capacity of humankind has stayed fairly constant during recent decades (actually during the recent millennia). It is reasonable to assume that people "think" and "speak" approximately the same amount (e.g., 10^{10} people times a constant number of words per day), and that their neural memory capacity has not significantly changed over the last two decades. In defense of this

assumption is the well-known fact that biological evolution is slow, especially in comparison to technological change. Studies have shown that an increasing part of our general communicational apparatus is increasingly being channeled through technologies. People report spending less time in face-to-face interaction since being connected to the Internet (WIP, 2010, p. 181). Therefore, the focus of understanding the role of information in social evolution points to the radical changes introduced by our technological information processing devices.

In the balance of Part I of this two-part article, we review decisions (b) and (c) in detail, while we review (d) (on unit of measurement) in Part II of this article. More details of each decision can be found in the almost 300-page online Supporting Appendix to this article.

Communication vs. Storage vs. Computation

We look at the question of technologically mediated information from the point of view of the various kinds of technologies. In agreement with mainstream studies of technologies we can view technologies as solutions to prototypical problems (e.g., see Arthur, 2009; Dosi, 1984; Sahal, 1985). ICT provides technological solutions to three kinds of prototypical problems: (1) how to transmit information between points A and B (the transmission of information through space, or "communication"); (2) how to transmit information from time 1 to 2 (the transmission of information through time, or "storage"); and (3) how to transform information ("computation") (see Figure 3; also Hilbert & Cairo, 2009).

We cover 60 types of technologies (see Figure 3). This compares to 25 covered by Bohn and Short (2009), 29 covered by Lyman et al. (2003), and 33 tracked by Gantz et al. (2008).

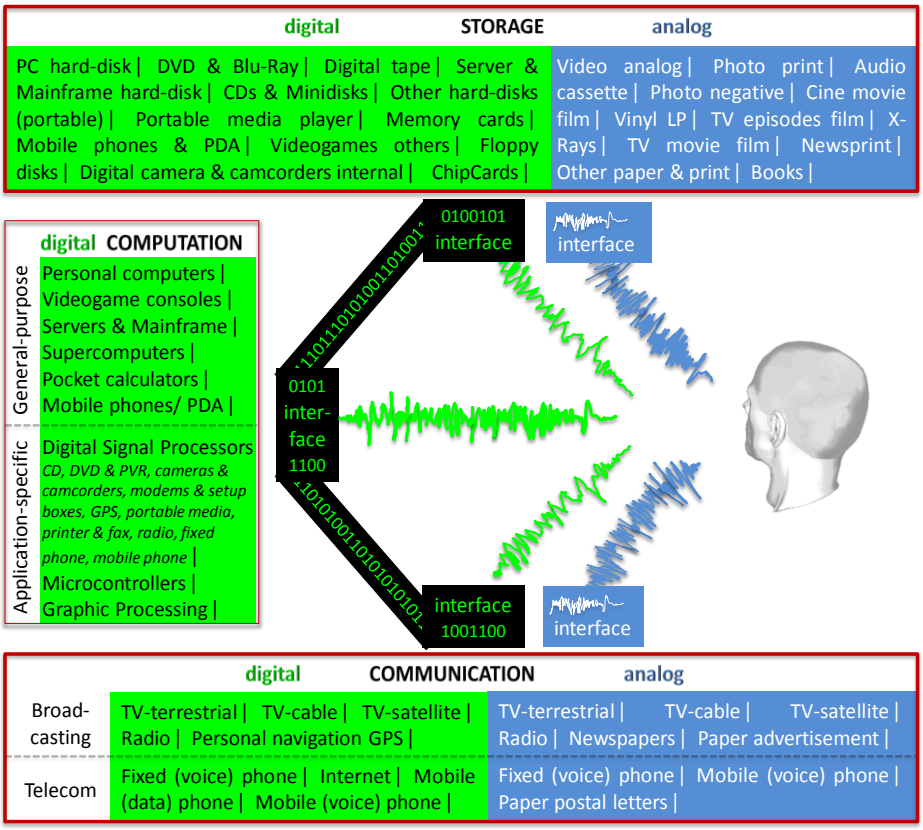

Figure 3. Three basic information operations and their most prominent technologies.
Source: Hilbert and López, 2011.

- We define *storage capacity* as the *capacity to maintain information over a considerable amount of time* for explicit later retrieval.

- We define *communication capacity* as the capacity to *receive or send information, while being transmitted over a considerable distance* outside the local area. Here we distinguish between broadcasting (roughly speaking, "one-way transmission") and telecommunication ("two-way transmission").

- *Computation capacity* is defined as the capacity to *meaningfully transform of information* according to a set of instructions, either from X into x (translation of symbols) or from X into Y (transformation of meaning).

Naturally, these definitions include a number of choices that we had to make and we provide more details to assure an unambiguous interpretation of these classes of information operations for our purposes. Let's start by taking a look at some of the grey zones among the three main groups. For example, newspapers and letters serve both communication and (temporal) storage functions. It would be justifiable to include them in either or in both categories (we opt for the latter). Books, videogame consoles, and DVDs serve as information storage, but could also be classified as communication, because one can send them from one place to another. Similarly, stacks of papers might be carried in the office from one room to the other or distributed at a meeting. This also communicates information. However, in this case we opt to only count the books, consoles, DVDs, and stacks of paper in storage, and do not include them in the world's communication capacity. The main reason here is not theoretical or conceptual, but mainly one of measurement feasibility. We do not have access to the statistics about how many books and DVDs are sent around over a considerable distance. Something similar accounts for information transmitted in closed local networks that do not make part of the open Internet (such as closed company LANs, private networks, or surveillance camera networks). The required statistics to include these local networks of communication do not exist, hence the aggregate "*over a considerable distance outside the local area*" in our definition.

Within the communication group, we further distinguish between broadcasting and telecom. This is useful for analytical purposes, mainly because broadcasting transmits many more bits than telecommunications. However, ICT convergence currently blurs this distinction. While traditionally broadcasting technologies distribute the same common content through exclusive downstream capacities, telecommunication technologies count with both up- and downstream channels, and the user can individually chose the content.[8] For example, digital television counts with a (quite small, but existent) upstream channel, which allows the user to send individual messages. Since until 2007 this upstream channel was not well used (besides some sporadic video-on-demand application) we decided to classify digital television as a broadcasting media, but this might not be as easily justifiable in the near future.

There are also grey lines between computation and the other functions. Thinking of computation in terms of Turing's (1937) famous setup, the manipulation of symbols in a universal computer takes place on a strip of storage tape according to a table of rules. The *communication* of information from the table of rules and the replacement of the information on the *storage* memory are normally carried out at high speed and according to a particular procedure (an algorithmic information recipe: "an algorithm is an ordered set of unambiguous, executable steps that defines a terminating process," (Brookshear, 2009, p. 205; for more on algorithms, see Cormen, Leiserson, Rivest, & Stein, 2003). While it would certainly be outside the scope of this article to go deeper into the intricate relatedness of storage (time), communication (space), and transformation of information in our fourth-dimensional space-time universe, the slippery repercussions of these relations become clear when thinking about the classification of

[8] We refer to the "downstream" and "upstream" capacity of communication, which refers to the flow of information, because we stress the flow of information, independent from its technological nature. In contrast, "download" and "upload" capacity refers more specifically to bringing data from a remote source to local storage and storing it there (Laplante, 1999), and "downlink" and "uplink" are mainly used to stress the wireless and mobile nature of the communication.

something like the supporting memory of RAM (random access memory). It would surely be justifiable to classify RAM as storage and/or computation. RAM memory acts like a temporary notepad that the computer's processing unit uses when making calculations. Since its primary function is not storage, but to assist computation, we decided not to include RAM in storage (hence the aggregate "over a considerable amount of time for explicit later retrieval" in our definition of *storage*). We assumed that RAM capacity is already measured by the MIPS indicator of the hardware computational capacity of a computer.

In this sense, for our inventory we took the methodological decision that the primary function of the information operation dictates the classification of what is included into the groups of storage, communication, or computation. This being said, it is important to point out that the cake could have been cut a different way. Bohn and Short (2009) look at "interactive versus passive," and one could also look at "video versus text," "wireless versus wired," "playback versus live," etc. If thoroughly justified, all of them simply take different perspectives and are equally valid.

Counting Devices vs. Expenditure

If we want to know how much information is handled by a group of technologies, we need to consider at least two basic indicators: one for the quantity of technologies out there, and another one for their performance. The most common way to quantify the technologies consists of an inventory of the number of devices. The next step consists of estimating the respective performance of each device. The sum of the product of both results is the total capacity. We use the term *performance* for the performance of an individual device, and *capacity* for the sum of the performance of all devices:

$$\sum_{over\ all\ i} [\text{devices of type i}] * [\text{performance per device of type i}] = capacity$$

An alternative way has been proposed by Short et al. (2011). They estimate the quantity of information effectively processed by enterprise servers by implicitly assuming that companies spend dollars for server capacity in the most efficient way for their specific requirements. Instead of counting the number of devices (servers), they measure the amount of dollars spent to purchase those devices, and instead of measuring the performance per device, they measure the performance per dollar spent (according to some benchmark tests). While the previous approach requires information about which kind of device provides which kind of performance, this alternative approach requires information about which kind of spending refers to which kind of performance:

$$\sum_{over\ all\ i} [\$\ \text{spent on process i}] * [\text{performance per \$ of process i}] = capacity$$

Once again, the choice of one or the other method is most likely not defined by theoretical or conceptual considerations, but is dictated by the availability of records, statistics, and other sources.

Technological Capacity vs. Consumption

Some studies focus on information consumption (Bohn & Short, 2009; Neuman et al., (this Special Section); Pool, 1983), and others, like ours, focus on the technological capacity (Gantz et al., 2008; Hilbert and López, 2011; Lesk, 1997). It is important to point out that the latter (capacity) is a required intermediary step to estimate the first (consumption). To estimate information consumption, one needs to have an understanding of the underlying technological capacity to process information. It is then straightforward to obtain the consumption by only considering a certain percentage of the installed capacity (weighted by the period during which information is effectively consumed).

One benefit of measuring consumption is that it tells us more about the social level of technology adoption. The behavior of the user moves into the center, independent from the respective technology. For example, Bohn and Short (2009) do not discriminate if the consumed information is received by a unidirectional real-time TV broadcasting network, a bidirectional telecommunicating mobile phone, or a DVD or videogame console that retrieves the information from a storage device. This might be justified for some purposes, but it blurs distinctions that might be relevant for other research questions. We start to see how the specific question on the researcher's mind defines the choice of indicator.

A downside of the consumption metric is that the estimation of consumption bears additional sources of error. Media consumption studies are resource intensive and are either based on small-scale sampling (e.g., WIP, 2010) or are only available for selected countries (e.g., BLS, 2010), which do not allow for international extrapolations.

But even where statistics are available, it is often not easy to measure information consumption unambiguously: Is every minute of media consumption equally intense? Does media consumption include peeking at a TV screen in a waiting room or at a shopping mall window? What about a radio running in the background? When staring at a computer monitor with an Internet Web page during a three-hour session, how many minutes of it are media consumption and how many minutes are thoughts going astray without consuming information? Is there a difference in information consumption intensity when halfheartedly staring at a TV screen versus playing an interactive videogame? How to count multitasking?[9] These kinds of methodological challenges can lead to ambiguity and discrepancies between sources. Let's have a closer look at the issue.

Installed Capacity Potential vs. Effective Usage Capacity vs. Consumption vs. Attention

To avoid losing ourselves in abstractness, let's walk through the case of communication to dig into these differences. The most straightforward metric to measure is what we call "installed capacity," or "installed capacity potential" (or "installed bandwidth potential" in the case of communication). It follows

[9] One solution is to sum different media usage times, such as was done by Bohn and Short (2009), which means that a single person could consume more than 24 hours of media per day (see Bohn & Short, this Special Section).

the above formula of the sum of the product of the number of devices and their reported capacity. Figure 4 depicts this capacity with the striped squares. End-user A, for example, maintains a DSL broadband subscription of 1 Mbps, while end-user B installed a cable-modem subscription that promises 2 Mbps, etc. The installed bandwidth potential of all four end-users is [1 Mbps**+**2 Mbps**+**3 Mbps**+**4 Mbps] = 10 Mbps.

We could now have a look at time-budget studies and see how long people report to interact with the technology on a regular basis. As shown with the winged brackets in Figure 4, end-user A reports interacting with the device for three hours per day, while end-user B uses two hours per day. Adjusting the installed bandwidth capacity with these reported usage times provides what some studies refer to as "consumption" (Bohn & Short, 2009; Neuman et al., this Special Section). Calculating the corresponding minutes of usage per day and converting it to an average usage per second (1 day=24 hours/day=2460 minutes/hour = 24*60*60 seconds/minute), we end up with an average of 0.5 Mbps, which is quite different from the above 10 Mbps.

Something additional that can be done is to apply some kind of "fudge factor" to the time-budget studies, which accounts for the fact that not every minute of interaction with a certain device has the same intensity and considers a certain "percentage of inattention" (Pool, 1983, p. 610). We could therefore distinguish between a "gross" and "net" rate of consumption. However, in reality, data sources on the question of attention are scare, specialized, or ambiguous (see the questions at the end of prior section). In principle, this would not change the logic outlined in Figure 4. Instead of considering the number of minutes that an individual interacts with the technology, we would merely consider those minutes to which the user pays a certain level of attention.

An alternative way is to measure the "effective capacity," or the "effective usage capacity." The distinction between "installed" and "effective usage" capacity is similar to what the Orbicom Infostate report (one of the pioneering international ICT indexes[10]) referred to as "Info-intensity" and "Info-use" (Sciadas, 2005). This measure does not require time-budget studies or inattention "fudge factors," but data about the effective traffic flow. The effective usage capacity tracks only those bits that are effectively communicated through the installed capacity. One could also say that the "effectively used" bits are "produced" in a given moment (Short et al., 2011 talk about information "production"), which can be different from saying that they are "consumed" by a user or not. For broadcast technologies (such as TVs) the "gross consumption" and the "effective usage capacity" are essentially the same. This is the reason why studies from the pre-Internet age did not have to worry about this distinction. However, considering modern digital telecommunication or computing power, it is clear that the "effective traffic" of a modem or the "effective duty cycles" of a computer do not equal the amount of time the user reports interacting with the device, or the time the device is "turned on." Therefore, we distinguish between installed and effective capacities.

To obtain the effective usage capacity we have to multiply the installed capacity not by the number of minutes reported by the individual for "interaction with the device" (through time-budget studies), but by the percentage by which the device is actually processing information. Continuing with

[10] For more on ICT indexes, see Minges, 2005.

our communication example of Figure 4, end-user A uses his 1 Mbps subscription 1% of the time, while end-user C deploys her 3 Mbps subscription 7% of the time, etc. The rest of the time, their monitors might be open and they might read emails or work on something else, but no information is communicated through the channel. As a result, we obtain a total capacity of 0.3 Mbps (see middle-squares in Figure 4). In practice we could obtain the numbers for the effective capacity by looking at the traffic in the backbone infrastructure (which is what we did for the estimations in Hilbert and López, 2011). This, however, once again excludes traffic that goes through local and private networks. To obtain these numbers one would need to get data from the traffic flows of individual modems or other subscriptions.

It should be clear by now that the reason why the three methods lead to three different numbers (10 Mbps, 0.5 Mbps, 0.3 Mbps) is that they measure different things.[11] What to measure is a question of interest, not one of validity. If the researcher is more interested in consumer behavior and media attention, consumption measures (with or without the "inattention fudge factor") might be the most adequate choice. If the researcher is more interested in the world's technological capacity, the choice will be different. If the researcher is not interested in estimating the absolute amount of information, but merely the distribution of capacity among different users, and is satisfied with the assumption that consumption patterns and attention percentages are constant among users (see, for example, Hilbert, 2011a), it might be sufficient to work with the "installed bandwidth potential." If the researcher was interested in the differences between the impact of TV and Internet traffic flow, a mix between "effective usage capacity" and "consumption (adjusted for inattention)" might be required to get to the bottom of this question. We can now clearly see that the question on the mind of the researcher will heavily influence the outcome of the inventory.

[11] In principle, the effective capacity could also be larger than the consumed information (for example, when users download movies they never watch), but the installed bandwidth potential will always be largest.

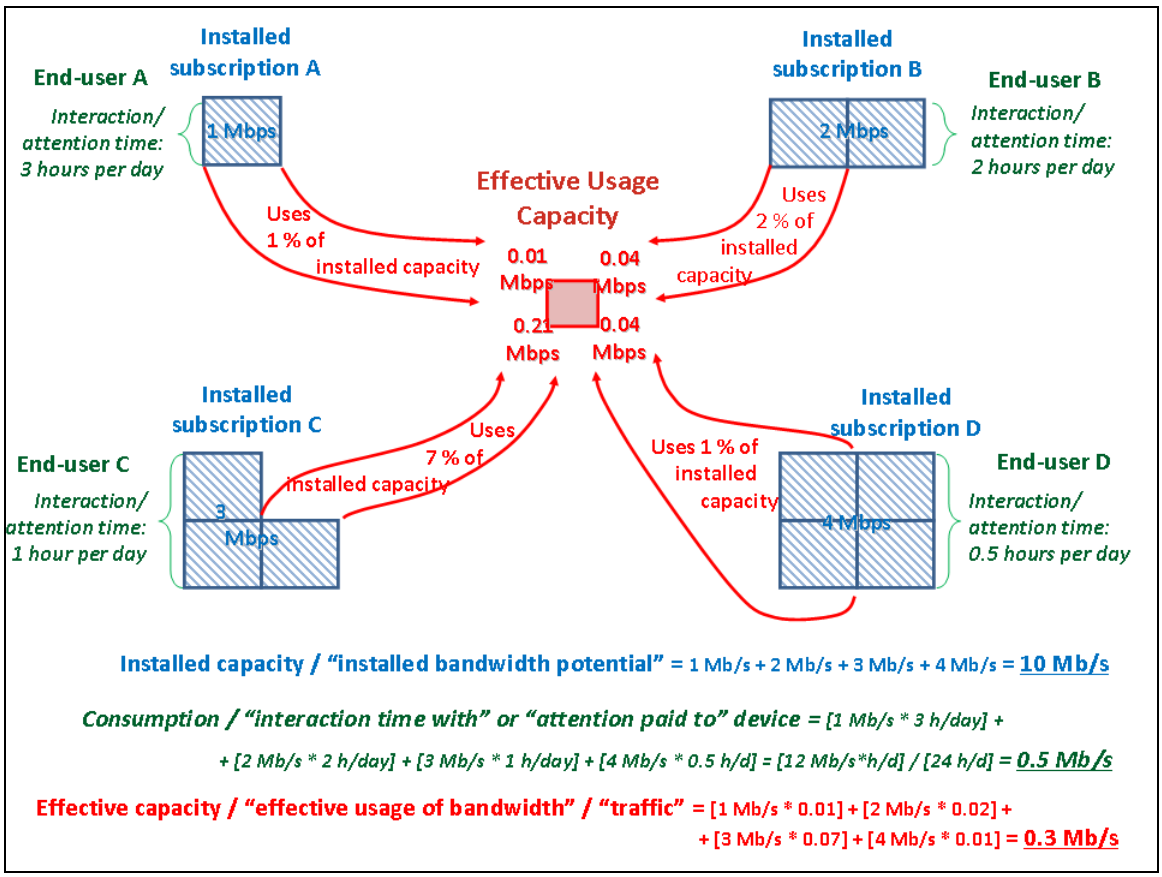

Figure 4. Installed capacity potential vs. effective usage capacity vs. consumption.
Source. Author's elaboration.

In our exercises we work with the "installed capacity potential" and "effective usage capacity." In the case of storage, we do not have reliable statistics that would reveal the effective usage of the technology, and we therefore report the "installed capacity potential." In other words, we do not know how much of a hard disk is effectively filled and how much is empty, and therefore simply estimate how much capacity is installed. We therefore estimate the theoretical maximum that could be used in a given point of time ("installed capacity"). In case more reliable statics on effective usage become available, it would be forthright to apply these percentages to convert installed storage capacity into effective (usage)

capacity. With the help of additional time-budget and attention studies, it would also be straightforward to convert them into the respective consumption estimates.

For the case of communication, our results highlight the difference between the "installed" and the "effective usage" capacities. In contrast to telecommunications networks, broadcast networks do not compete for resources to transmit information.[12] All broadcast receivers could receive information for 24 hour per day (this could have communicated 15.9 zettabytes in 2007). Effectively, however, our estimates show that the effective capacity was at 1.9 zettabytes, which results from the fact that the average broadcasting receiver runs for only three hours of the possible 24 hours per day (1.9/15.9). Telecommunications networks, on the contrary, would collapse if all subscriptions demanded to communicate at the same moment (similar to what happens on New Year's Eve to the mobile phone network). In a telecommunications network, all users compete for a limited infrastructure, and it works because each of them only uses a small part of it during short periods. According to our estimates, if all Internet subscriptions would run at their promised potential 24 hours per day, the world would need a network infrastructure that could carry 13.6 zettabytes in 2007. Going with the "consumption-time-budget studies" figures, people report to "use" or "consume" the Internet for 1.6 hours per day on average, which reduces the required network infrastructure to 907 exabytes (13,600*1.6/24) of effective gross usage.

This is still an overestimation since it refers to the time people report to interact with the technology (e.g., sit in front of the monitor) and not the amount of traffic that actually flows. (We refer to the mere interaction with the device as "gross" and the active flow/processing of information as "net" effective capacities simply to give them names in order to distinguish them.) Comparing the numbers reported by Odlyzko (2010) and Cisco Systems (2008) about the installed Internet backbone infrastructure (some 68 exabytes), it turns out that the average user only uses its promised full bandwidth for effectively nine minutes per day. During the remaining 87 minutes of the session, the screen is open, but no telecommunication takes place through the modem (see also Hilbert, 2011a). It is important to point out that the numbers of Odlyzko (2010) and Cisco Systems (2008) only take inventory of the traffic in the larger backbone network,[13] and do not include local traffic that never hits the backbone (for a brief discussion, see Bohn & Short, 2009). Since these local area data are unavailable, we do not include them in our global estimates of the "effective usage capacity," hence, our definition of telecommunication in terms of information "transmitted over a considerable distance outside the local area."

---

[12] On technical grounds the distinction is that broadcasting transmits one single signal to all receivers, while in the case of telecommunications, every user counts with an individually usable channel. The user-defined telecom channel competes for a shared bandwidth (at the backbone level), while the common broadcast signal is non-exclusive.

[13] Assuming that telecom investors make (more or less) sensible calculations that aim at balancing the supply and (shared) demand for bandwidth, the installed infrastructure allows us to approximate the effective capacity, plus an additional margin of error due to market inefficiencies (e.g., over- or sub-installation of infrastructure, see Odlyzko, 2003, 2008). On the contrary, in the case of paper-based communication, the statistic of the effective usage capacity is much more readily available than the "installed potential capacity": How many letters could a mail carrier maximally deliver?

In the case of general-purpose computation, we elaborate both the installed capacity and the "gross effective capacity." This means that we multiply the installed capacity with the time users report to "interact" with the machine, which does not reveal the effective "net" capacity during which the computer is running at 100% of its processing capacity (does not count effective "duty-cycles" of the effective "net" capacity). In an aggregation of different technologies this skews the results to favor technologies with long periods of "computation-idle" usage and "low duty cycles" (such as mobile phones) versus technologies with high-intensity computation sessions (such as supercomputers). However, we do not not have statistics on the "effective net usage time" or "effectively executed duty cycles per day" of all computing devices included in our inventory.

Figure 5 shows the results for "installed" versus "effective gross usage" capacities for 2007. It shows that, in relative terms, enterprise servers and supercomputers gain importance when transforming "installed capacity" into "effective gross capacity," because they run more hours per day than personal computers or videogame consoles. We estimate that enterprise servers are (gross) used a third of the time, which agrees with the estimates from Short et al. (2011), who carry out a more detailed inventory for enterprise servers than we do. Overall, our estimates report that over the last two decades the effective gross capacity of general-purpose computers has been between 6–9% of the installed capacity. In other words, humankind has (at least) between 10–20 times more computational power in its hands than it actually uses (more if considering that "gross effective capacity" almost never demands 100% of the available "net effective capacity").

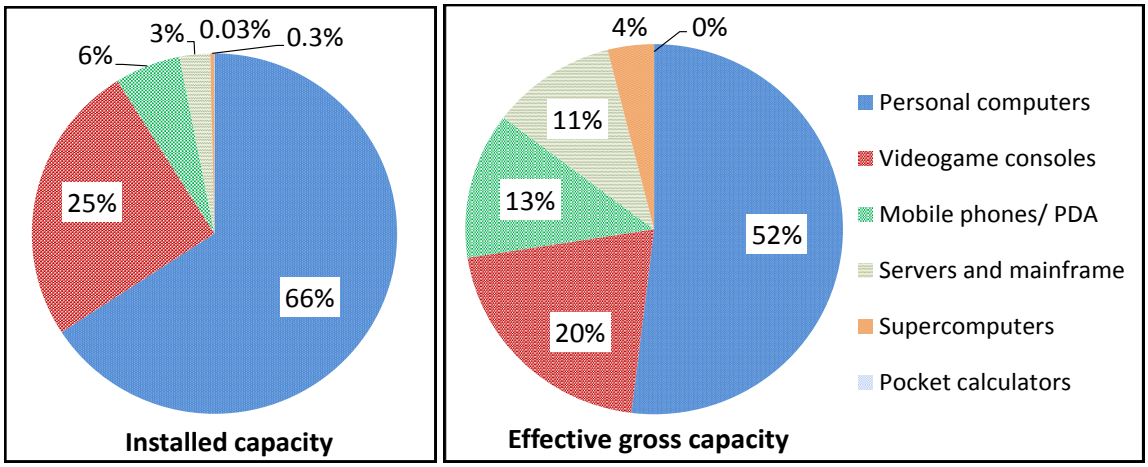

Figure 5. General-purpose computation 2007:
Installed capacity versus effective gross capacity.
Source. Hilbert and López, 2011, also Appendix.

Supply (Sent) vs. Demand (Received)

Methodologically intertwined with the previous distinctions is the question of accounting for supplied or sent messages versus demanded or received messages. Lyman et al. (2000) and Lyman et al. (2003) define communication in terms of the sent messages (see also Bounie, 2003). On the contrary, in our inventory we follow Shannon's mathematical theory of communication (1948), according to which communication consists of the reduction of uncertainty[14] (see also Cover & Thomas, 2006; Massey, 1998). In this interpretation, a message that is sent over a channel and never picked up and decoded does not constitute information (since it does not reduce any uncertainty). The distinction is reminiscent of the old question: If a tree that falls in the woods, does it make a sound if nobody is there to hear it? If nobody is there to pick up the signals, the falling tree merely produces a shift in the movement of molecules through the air. Some say that only if somebody decodes this movement of molecules and translates it into the cognitive perception of a sound (or any other kind of meaningful signal, such as a visual diagram or color) does communication take place. However, one might just as well be interested in how many TV signals are broadcast, independently from the fact that somebody picks them up or not. Neuman et al. (this Special Section) directly compare the supply and demand of information and find an increasing discrepancy between them, due to the explosion of information supply. Once again, the question on the researcher's mind guides the methodological choice.

Technological Capacity vs. Unique Information

Lyman et al. (2000), Lyman et al. (2003), and Bounie (2003) aimed at an inventory of the amount of new, unique, and original information produced, which differs from our estimate of the installed technological capacity or of an information consumption measure. Lyman et al. (2000) and Lyman et al. (2003) estimated that the world produced between 2–3 exabytes of uniquely new information in 1999 and 5.4 exabytes in 2002. We estimate that the world's technological capacity of storing and of communicating optimally compressed information in 2000 reached 54 and 1,150 exabytes, respectively, which is roughly between 15–300 times as much, including copies. This also leads to important differences in the shares different technologies contribute to the total. Film captures a much larger share in Lyman et al. (2000) and Lyman et al. (2003) than in ours (7.7% vs. 0.64% in 2000) and optical storage a much smaller share (0.002% vs. 10.5%), because film is mainly used to capture original information (i.e., for photographs, movies, or TV), while optical storage, such as CD-ROM and DVD, are mainly used as backup storage or as distribution medium.[15]

Their focus on unique information leads Lyman et al. (2000) and Lyman et al. (2003) to conclude that "most of the total volume of new information flows is derived from the volume of voice telephone traffic, most of which is unique content" (97%). Television content, which broadcasts the same

[14] More precisely, the "mutual information" between X and Y is the reduction in the uncertainty of X, due to what Y already knows: $I(X;Y) = H(Y) - H(Y|X)$.

[15] In both of our estimations, magnetic storage makes up the majority in 2000 (92% in Berkeley's estimation and 87% in ours), while paper contributes a minor role (0.03% of unique information from the Berkeley estimate and 1.08% if one considers our assumptions of accumulation).

information over many channels, accounts for the lion's share of our communication total (between 93–96% of total effective capacity). For 2002, Lyman et al. (2003) estimated that 21,264 TV stations worldwide produced about 31 million hours of original programming (69 petabytes). We estimate that 1.6 billion analog and digital TV receivers effectively displayed 1,113,000 optimally compressed petabytes in the year 2000. This difference also affects the relation between communication and storage. Lyman et al. (2003) found that information flows through electronic channels (telephone, radio, TV, and the Internet) contained around 3.5 times more original information in 2002, than what was recorded in storage media. We estimate that there is 21 times more communication than storage capacity in 2000, and seven times more in 2007.

Once again, the question on the researcher's mind determines which indicator to use. As with consumption, the concept of "unique information" is in practice a more advanced step, which often builds on the calculation of the technological capacity. Based on the capacity, it is straightforward to consider only a certain percentage of the capacity as "unique." However, finding this percentage is not easy, since information is never totally "original" and "unique," but always a recombination of an existing alphabet. For example, is a "new" rock music song really unique or just a recombination of a series of existing G, C, and D chords? Does resampling of the order of songs on a playlist constitute new and unique information or is it still the same as before? What if a song is cut short and a new song starts early? From an information theoretic perspective, information is always simply the recombination of symbols that are taken from a previously defined and existing alphabet. Text, for example, is the recombination of certain existing words on a page. When does a text say something unique and new, and when is it the same as has been said before? The grey zone seems large and the definition of what to count as unique or duplicate will depend on the specific definition elaborated by the researcher. Without a viable working definition of what "new and unique information" is, it will be hard to measure it.

Bohn and Short (2009), Gantz et al. (2008), and our exercise therefore abandoned the lead of Lyman et al. (2000) and Lyman et al. (2003) in this regard, and does not distinguish between unique and duplicate information. This is appropriate, for example, if the researcher is not interested in the amount of "innovation" in information production, but merely in the usefulness of information. The pure fact that somebody or something stores or communicates a particular amount of information indicates that this amount of information must be useful for somebody or something (independent of the fact that it might be useful to another individual as well). In Part II of this article we propose a thought experiment that aims at unambiguously identifying unique information (see Box 2 of Part II), but unfortunately, it is merely a thought experiment.

Theoretical Focus vs. Measurement Feasibility

Before moving on to Part II of this article (which focuses on the unit of measurement: optimally compressed bits in our case), let's pause to underline two general insights from the preceding discussion. We have seen that two main determinants influence the final result of any information inventory (and usually of any data gathering exercise in general):

1. The question on the mind of the researcher
2. Data availability

While the first one is of a conceptual nature, the second one is brutally practical. Both often influence the methodological choices to a similar degree, and the result is (as always) a trade-off between the theoretically desirable and the practically feasible. Differences in results of the various inventories result from either one or the other factor: differences in the questions asked and the existence of data to tackle this question. In practice, there is often a trade-off between the two. As long as the basic scientific codex is followed (justification of choices, transparency in methodological decisions and sources, etc.), all of the results are equally valid in an academic sense and all contribute the best they can to the final goal of understanding the phenomena better.

This said, it is important to stress that a concrete research question always tackles a limited part of reality, and the chosen theoretical framework defines the bias by which indicators are defined to answer the chosen question. In other words, the selected theoretical framework defines which aspects of reality to blank out and on which to set the focus. This inevitably introduces a dynamic of self-confirmation, whereas indicators are the result and, at the same time, the supposed empirical testing ground for the validity of the chosen theory. Theories and indicators relate to each other like the famous chicken and the egg, and the measurement of information in society is no exception.

Appendix

We place large emphasis on transparency in outlining the methodological assumptions and sources on the basis of which we elaborated the presented estimates. More than arguing in favor of one specific number, we see the presented estimates as an approximation, which could certainly be improved (depending on the available resources). To facilitate future generations of research on this topic, 300 pages of Supporting Appendix that outline the details of the applied methodology, enlisting more than 1,100 distinct sources to include:

Supporting Appendix, Material, and Methods
A. Statistical Lessons Learned
B. Compression
C. Storage
D. Communication: incl. update for telecommunications (telephony and Internet) until 2010
E. Computation

This Supporting Appendix can be accessed at http://www.martinhilbert.net/WorldInfoCapacity.html

References

Arthur, W. B. (2009). *The nature of technology: What it is and how it evolves*. New York: Free Press.

Bär, N. (2011, February 11). El planeta, ahogado en información [The planet, choked on information]. *La Nacion*. Buenos Aires. Retrieved from http://www.lanacion.com.ar/1349045-el-planeta-ahogado-en-informacion

Bialik, C. (2011, February 19). The world's information explosion. *The Wall Street Journal*. Retrieved from http://online.wsj.com/article/SB10001424052748704900004576152384123140652.html

Bohn, R., & Short, J. (2009). *How much information? 2009 report on American consumers*. San Diego: Global Information Industry Center of University of California, San Diego. Retrieved from http://hmi.ucsd.edu/howmuchinfo.php

Bounie, D. (2003). *The international production and dissemination of information*. Special Project on The Economics of Knowledge, Autorità per le Garanzie nelle Comunicazioni. Paris: École Nationale Supérieure des Télécommunications (ENST). Retrieved from http://ses.telecom-paristech.fr/bounie/documents/Recherche/Annex.pdf

Brookshear, J. G. (2009). *Computer science: An overview* (10th ed.). Upper Saddle River, NJ: Addison Wesley.

Castells, M. (2009). *The rise of the network society: The information age: Economy, society, and culture* Vol. I (2nd ed.). West Sussex:mWiley-Blackwell.

Cisco Systems. (2008). *Global IP traffic forecast and methodology, 2006–2011* (White Paper). Retrieved from http://www.hbtf.org/files/cisco_IPforecast.pdf

Cloud Storage. (2010, December 30). *Cloud storage*. Retrieved from http://en.wikipedia.org/wiki/Cloud_storage

Cormen, T., Leiserson, C., Rivest, R., & Stein, C. (2003). *Introduction to algorithms* (2nd ed.). Boston: McGraw-Hill Science/Engineering/Math.

Coughlin, T. (2007). *Digital storage technology newsletters*. Coughlin Associates. Retrieved from http://www.tomcoughlin.com

Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, NJ: Wiley-Interscience.

de Luna, T. (2011, February). Necessidade de espaço para armazenar dados digitais é cada vez maior [Demand for digital data storage is ever increasing]. *Correio Braziliense*. Retrieved from http://www.correiobraziliense.com.br/app/noticia/tecnologia/2011/02/11/interna_tecnologia,237233/necessidade-de-espaco-para-armazenar-dados-digitais-e-cada-vez-maior.shtml

Dosi, G. (1984). *Technical change and industrial transformation: Theory and an application to the semiconductor industry*. London: Palgrave Macmillan.

Etro, F. (2009). The economic impact of cloud computing on business creation, employment and output in Europe. An application of the endogenous market structures approach to a GPT innovation. *Review of Business and Economics*, *LIV*(2), 179–208.

Faostat. (2010). *ForesSTAT*. United Nations Food and Agriculture Organization. Retrieved from http://faostat.fao.org

Gantz, J., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W., et al. (2008). *The diverse and exploding digital universe: An updated forecast of worldwide information growth through 2011*. Framingham, MA: IDC (International Data Corporation) sponsored by EMC. Retrieved from http://www.emc.com/leadership/digital-universe/expanding-digital-universe.htm

Greenemeier, L. (2011, February 10). Down in the data dumps: Researchers inventory a world of information. *Scientific American*. Retrieved from http://www.scientificamerican.com/article.cfm?id=information-inventory

Hilbert, M. (2011a). *Mapping the dimensions and characteristics of the world's technological communication capacity during the period of digitization*. Presented at the 9th World Telecommunication/ICT Indicators Meeting, Mauritius: International Telecommunication Union (ITU). Retrieved from http://www.itu.int/ITU-D/ict/wtim11/documents/inf/015INF-E.pdf

Hilbert, M. (2011b). *That giant sifting sound*. Video presentation at *The Economist* Ideas Economy: Information Summit. The Economist, proof-inc. Retrieved from http://ideas.economist.com/video/giant-sifting-sound-0

Hilbert, M., & Cairo, O. (Eds.). (2009). *Quo Vaids, Tecnologia de la informacion y de las comunicaciones* [Quo Vadis information and communication technology]. Bogota: Mayol Ediciones. Retrieved from http://www.eclac.org/SocInfo

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60–65. doi:10.1126/science.1200970

International Telecommunication Union (ITU). (2010). *World telecommunication/ICT indicators database*. Geneva: ITU. Retrieved from http://www.itu.int/ITU-D/ict/statistics

IFPI. (2007). *The recording industry world sales 1995–2004*. International Federation of the Phonographic Industry. Retrieved from http://www.ifpi.org/content/section_statistics/index.html

Ito, Y. (1981). The Johoka Shakai approach to the study of communication in Japan. In C. Wilhoit & H. de Bock (Eds.), *Mass Communication Review Yearbook* (Vol. 2, pp. 671–698). Beverly Hills, CA: SAGE Publications.

Japanese Recording-Media Industries Association (JRIA). (2007). *Press releases*. JRIA. Tokyo, Japan.

Laplante, P. A. (1999). Electrical engineering dictionary, CRCnetBASE 1999. Heidelberg: Springer.

Lebwohl, B. (2011, February 10). *Martin Hilbert: All human information, stored on CD, would reach beyond the moon*. Earthsky: A clear voice for science, human world. Retrieved from http://earthsky.org/human-world/martin-hilbert-all-human-information-stored-on-cd-would-reach-beyond-the-moon

Lesk, M. (1997). *How much information is there in the world?* lesk.com. Retrieved from http://www.lesk.com/mlesk/ksg97/ksg.html

Longbottom, R. (2006). *Computer speed claims 1980 to 1996*. Roy Longbottom's PC Benchmark collection. Retrieved from http://www.roylongbottom.org.uk/mips.htm

López, P., & Hilbert, M. (2011). Supporting online material: The world's technological capacity to store, communicate, and compute information. *Science*, *332*(6025), 60–65. doi:10.1126/science.1200970

Lyman, P., Varian, H. R., Dunn, J., Strygin, A., & Swearingen, K. (2000). *How much information 2000*. University of California, Berkeley. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info

Lyman, P., Varian, H., Swearingen, K., Charles, P., Good, N., Jordan, L., et al. (2003). *How much information? 2003*. University of California, Berkeley. Retrieved from http://www2.sims.berkeley.edu/research/projects/how-much-info-2003

Mansell, R. (2009). *The information society*. London: Routledge.

Massey, J. (1998). *Applied digital information theory: Lecture notes by Prof. em. J. L. Massey*. Swiss Federal Institute of Technology. Retrieved from http://www.isiweb.ee.ethz.ch/archive/massey_scr

McAllister, H. (1997). Grains of sand on the world's beaches. *Suremath, University of Hawai'i system*. Retrieved from http://www.hawaii.edu/suremath/jsand.html

McCallum, J. (2002). Price-performance of computer technology. In V. G. Oklobdzija (Ed.), *The computer engineering handbook* (pp. 136–153). Boca Raton, Florida:CRC Press.

Meyer, D. T., Shamma, M., Wires, J., Zhang, Q., Hutchinson, N. C., & Warfield, A. (2010). Fast and cautious evolution of cloud storage. Presented at the *HotStorage '10*, Boston: USENIX: The Advanced Computing Systems Association. Retrieved from http://www.usenix.org/events/hotstorage10/tech/full_papers/Meyer.pdf

Minges, M. (2005). *Evaluation of e-readiness indices in Latin America and the Caribbean*. Santiago: United Nations ECLAC. Retrieved from http://www.cepal.org/SocInfo/publicaciones

Odlyzko, A. (2003). Internet traffic growth: Sources and implications. In B. B. Dingel, W. Weiershausen, A. K. Dutta, & K.-I. Sato (Eds.), Optical transmission systems and equipment for WDM networking II, Proceedings of the The International Society for Optical Engineering (SPIE) (pp. 1–15). Retrieved from http://adsabs.harvard.edu/abs/2003SPIE.5247.....D

Odlyzko, A. (2008, February 25). Threats to the Internet: Too much or too little growth? *Internet Evolution*. Retrieved from http://www.internetevolution.com/author.asp?section_id=592&doc_id=146747&

Odlyzko, A. (2010). *Minnesota Internet traffic studies (MINTS)*. University of Minnesota. Retrieved from http://www.dtc.umn.edu/mints

Pool, I. de S. (1983). Tracking the flow of information. *Science*, *221*(4611), 609–613. doi:10.1126/science.221.4611.609

Porter, J. (2005). *Disk/trend reports 1977–1999*. Retrieved from http://www.disktrend.com

Sahal, D. (1985). Technological guideposts and innovation avenues. *Research Policy*, *14*(2), 61–82. doi:10.1016/0048-7333(85)90015-0

Schrader, C. (2011, February 10). Explosion des cyberspace: Datenwachstum der digitalisierten welt [Explosion of cyberspace: Data growth of the digitalized world]. *Sueddeutsche Zeitung*. Retrieved from http://www.sueddeutsche.de/digital/datenwachstum-der-digitalisierten-welt-explosion-des-cyberspace-1.1058394

Sciadas, G. (2005, November). *From the digital divide to digital opportunities*. Canada, Montreal: ORBICOM.

Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27,* 379–423, 623–656. doi:10.1145/584091.584093

Short, J., Bohn, R., & Baru, C. (2011). *How much information? 2010 report on enterprise server information*. San Diego: Global Information Industry Center at the School of International Relations and Pacific Studies, UC-San Diego. Retrieved from http://hmi.ucsd.edu/howmuchinfo_research_report_consum_2010.php

Stewart, J. (2011, February 11). Global data storage calculated at 295 exabytes. *BBC News*. Retrieved from http://www.bbc.co.uk/news/technology-12419672

Technica, A. (2011, February 11). World's total CPU power: One human brain. *Wired*, *Wired Science*. Retrieved from http://www.wired.com/wiredscience/2011/02/world-computer-data

TOP500. (2009). *TOP500 list releases*. TOP500 supercomputer sites. Retrieved from http://www.top500.org/lists

Turing, A. M. (1937). On computable numbers, with an application to the entscheidungsproblem. *Proceedings of the London Mathematical Society*, *s2-42*(1), 230–265. doi:10.1112/plms/s2-42.1.230

Universal Postal Union (UPU). (2007). *Postal statistics*. UPU. Retrieved from http://www.upu.int/en/resources/postal-statistics

U.S. Bureau of Labor Statistics (BLS). (2010). American Time Use Survey (ATUS). Retrieved from http://www.bls.gov/tus

Vastag, B. (2011, February 10). Exabytes: Documenting the 'digital age' and huge growth in computing capacity. *The Washington Post*. Retrieved from http://www.washingtonpost.com/wp-dyn/content/article/2011/02/10/AR2011021004916.html

World Internet Project (WIP). (2010). *World Internet international report 2010*. Los Angeles: USC Annenberg School Center for the Digital Future.