

How Much Information is There in the “Information Society”?

Martin Hilbert

finally published as

Hilbert, M. (2012). How much information is there in the “information society”?
Significance, 9(4), 8–12.

doi:10.1111/j.1740-9713.2012.00584.x

<http://www.significancemagazine.org>

<http://onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2012.00584.x/abstract>

We know our brave new world is being transformed by data. But how much more of it is there than before? **Martin Hilbert** explores, and finds huge untapped capacity to process information.

During recent decades the world was swamped with technologies that enable us to store, communicate and compute information in an unprecedented order of magnitude. Social scientists from all disciplines have been relentless in pointing out that this has led us from the industrial to the information society, whereas information moved into the center of social, economic, political and cultural organization and development. And indeed, we can see the effects of digital information and communication technologies (ICTs) everywhere. We know that the digitization of information stocks and flows has been and still is the driver of economic productivity, the lynchpin for the introduction of transparency and efficiency gains in governmental bureaucracies, the key for cost savings in health reforms, leading in the modernization of education, and that the mobile phone even changes the way families coordinate and the internet how young people date and fall in love.

Despite the omnipresence of the information age, there is a surprising lack of information about the amount of information in the information society. This is in stark contrast with the fact that humankind's desire to quantify information has a long history. Aristotle's student Demetrius (367 BC–ca. 283 BC) was asked to organize the Library of Alexandria in order to quantify “how many thousand books

are there". The inventor of the bit, and therefore the intellectual father of the digital age, Claude E. Shannon, estimated in 1949 that the largest information stockpile he could think of contained some 12,500 MB, the U.S. Library of Congress.

Together with my coauthor Priscila López, we took up the much celebrated theories of Shannon and his colleagues and used them to create a methodology that allowed us to estimate the world's technological capacity to store, communicate and compute information over the period from 1986 to 2007/2010.¹ Over five years we gathered more than 1,100 different sources and took inventory of 60 analog and digital technologies, including books and newspapers, hard-disks and DVDs, internet subscription and mobile phones, vinyl records and audio-cassettes, videogame consoles and pocket calculators, and even every little chip on the back of your credit card. Only to list our sources and explain our methodological choices we had to produce nearly 300 pages of methodological notes.² But we learned several interesting things that we didn't know before.

We started by reconfirming that the quantification of technologically mediated information is so interesting right now precisely because it changed so much over recent decades. Since biological evolution is notoriously slow, it cannot reasonably be expected that human kinds' biological information processing capacity has changed more than population growth in recent decades (1 % - 1.5 % per year). In contrast, the world's technological capacity to store and telecommunicate information has grown with a compound annual growth rate of 25 – 30 % during the period from 1986 to 2007 (roughly five times faster than economic growth during the same period), and human kinds technological capacity to compute information has even grown between 60 % - 85 % annually (more than 10 times faster than our economic capacities).³ These are orders of magnitude that blow social scientists out of the tranquil waters of single digit rates of change they usually navigate.

The amount of information received through one-way broadcast networks was the slowest growing information operation. It "merely" quadruplet during the two decades from 1986 to 2007, from 432 exabytes, to 1.9 zettabytes. A zettabyte is a number with 21 zeros: a kilobyte is equivalent to '000, a megabyte to '000,000, a gigabyte to '000,000,000, and then it goes tera-, peta-, exa-, zetta-, and eventually you will start hearing about yottabytes (10^{24}). If we would like to transmit the amount of broadcasted information with the help of newspapers, we would have had to deliver 55 newspapers per person per day in 1986, and 175 newspapers per person per day by 2007.⁴ Surprisingly, the world's effective capacity to exchange information through our omnipresent two-way telecommunications networks (such as the internet and telephony) was and still is happening at a much smaller scale (see Figure 1). In 1986, the amount of

telecommunication was the informational equivalent to 2 newspaper pages per person per day. Notwithstanding this low starting level, since then our telecommunications capacity has grown four times faster than our broadcast capacity, reaching the informational equivalent of some 20 entire newspapers per person per day by 2010.⁴

ICTs do not only help us to transmit information through space (communication), but also through time (storage). Our global technological memory has roughly doubled every three years over recent decades, from less than 3 exabytes in 1986, to about 300 in 2007. If we would have liked to store this on double printed book paper, we could have covered every square centimeter of the world's landmasses with one sheet of paper in 1986, with one layer of books by 2007, and by 2010 with two layers of books.⁴ Our numbers also indicate that the year 2002 marked the beginning of the digital age, since this year marked the date where human kind started to store more information on digital, than on analog storage devices. This transition happened at the blink of an eye in historical terms. While merely 1 % of the world's capacity to store information was in digital format in 1986, our digital memory represented 25% of the total in the year 2000, and exploded to 97% of the world's storage capacity by 2007.

It is interesting to observe that the kind of content has not changed significantly since the analog age: despite the general perception that the digital age is synonymous with the proliferation of media-rich audio and videos, we find that text and still images capture a larger share of the world's technological memories than before the digital age.⁵ In the early 1990s, video represented more than 80 % of the world's information stock (mainly stored in analog VHS cassettes) and audio almost 15 % (audio cassettes and vinyl records). By 2007, the share of video in the world's storage devices decreased to 60 % and the share of audio to merely 5 %, while text increased from less than 1 % to a staggering 20 % (boosted by the vast amounts of alphanumerical content on internet servers, hard-disks and databases). The multi-media age actually turns out to be an alphanumeric text age, which is good news if you want to make life easy for search engines.

While our telecommunication and storage capacity has grown at roughly the same speed, the fact that we have traditionally already transmitted such vast amounts of information through broadcast networks results in the fact that we have always communicated much more information than we can possibly store. This is shown in Figure 1, in which the combined broadcast and telecommunication capacity is measured on the left-hand side (in the range of 10^{15} MB) and our storage capacity on the right-hand side (in the range of 10^{14} MB). However, since broadcasting did not grow as fast, this ratio is changing quickly and the share of the amount of communicated information that can be stored is catching up quickly.

While the entire world's effectively communicated information would have filled up our global storage capacity in roughly 2.2 days in 1986 (after that we would have started to delete content), it would have taken almost 8 weeks two decades later.

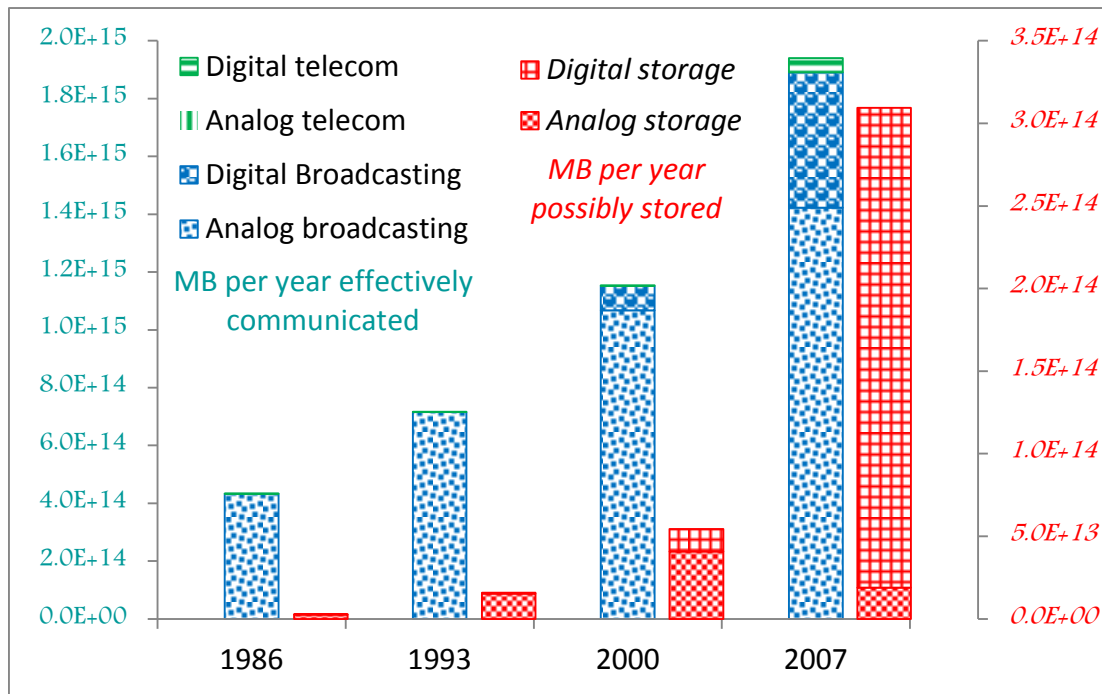


Figure 1. Installed capacity of storage and effective capacity of broadcasting and telecommunication, in optimally compressed megabytes (MB) per year for 1986, 1993, 2000, and 2007. Taken from⁶

The digital age is often taken to be synonymous with the telecommunicating internet and mobile phones, or large information storing server farms and databases. Surprisingly we found the fastest growing information operation has actually been none of these, but computation. A computer also stores and communicates information within its architecture, but very fast, on a very small scale and according to some deterministic 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: (humanly-guided) general-purpose computers (such as PCs, handheld devices, mainframe and servers, videogame consoles, etc.) and application-specific computers (embedded into electronic devices, household appliances or monitors, etc.). The world's general-purpose computing power has grown twice as fast as the world's storage and telecommunication capacity and the world's application-specific computers even three times as fast. This discrepancy in growth rate between general-purpose and application-specific computers has over the years also led to the fact that by the late 2000s application-specific computers have more than 20 times more

computing power than those technologies that we usually refer to as computers (of the general-purpose kind). This does not mean that the technological capacity of general-purpose computers is small. If some 2,200 people would execute manual calculations for a period that extends from the Big Bang until 2007, they could execute as many instructions as our general-purpose computers can execute in only one second.⁴

The fact that computation grows much faster than the growth of information is good news for those who are worried about the information overload. Since natural evolution is too slow to boost our biological cognitive abilities in the foreseeable future, the only option we have left to make sense of all the data is to fight fire with fire: using our own technological devices (i.e. artificially intelligent computers) to sift through the vast amounts of information delivered to us from our ever more powerful communication and storage technologies. The recently much heralded Big Data Analysis paradigm, and its leading disciples, such as Facebook, Amazon, and Google, have promised to make use of this trend and create value out of vast amounts of data through intelligent computational analysis.

Another interesting question relates to the main driver behind the information explosion: is it driven by more or by better technology? We constantly have more technology (growth in infrastructure) and we constantly have better technology (better hardware and better software). Using an analogy, for our communication and storage capacity the underlying logic is comparable with filling a growing number of buckets or tubes (infrastructure) of different sizes (hardware) with content of different levels of granularity (which refers to the software compression of the informational bits).⁷ The more fine-grained the compression of the content, the more fits, which is a detail that is often forgotten. Without considering compression one measures the hardware capacity of the technology⁸, but not its information content. Levels of compression have varied over the years, and vary among technologies. Some content is half-heartedly compressed (such as the 64 kbps of your fixed-line phone, which could be much further compressed without loss of quality), while other is almost optimally compressed (such as your mobile phone). In order to make the information content comparable, we normalize our estimates on the optimal level of compression. The uttermost level of compression has a special status in information theory, since it relates to the statistical nature of the source, which Shannon himself termed "entropy" (which is a more general form of thermodynamic entropy in physics).

It turns out that compression has been an important driver of the growth of our informational capacity. We found that the same amount of hardware can communicate and store more than 3 times more information in 2010 than 25 years earlier thanks to the more efficient software compression of information.⁹ In

general we noticed that the nature of technological change in the digital age has been changing over recent decades. In the late 1980s and early 1990s, our informational capacity has mainly been driven by the installation of more technological devices. We flooded the world with computers, optical disks and of course mobile phones (which is the fastest diffusing technology in the history of human kind, reaching 9 out of 10 people worldwide in less than two decades). However, in later years, the world's technological information capacity has been ever more driven by better, not merely more technology.¹⁰ Actually, the number of devices the average person can handle seems to reach a certain level of saturation. The average number of storage devices per person has stayed constant at 22-23 storage devices per capita since the year 2000 (including the number of books, hard-disks, mobile devices, CDs and DVDs you own, etc.). The average number of telecommunication devices does not seem to go beyond 2-3 devices per capita (roughly one or two phones and one internet subscription). This does not mean that our technological capacity to store or communicate information has come to a halt. Our technology simply always becomes better.

Overall, since the late 1980s, "better technology" has contributed more than twice as much to the growth in our telecommunication capacity than "more technology", and more than four times as much for the information explosion of stored information.¹⁰ This tendency will only intensify as we tend to update the performance of an existing number of devices and have stopped equipping us with additional ones. On the applied side, this has important consequences for the statistical practice of how to assess the digital age for public policies and private business strategies.

Over the recent decades, a well-functioning and institutionalized statistical apparatus has been created that involves technology regulatory authorities (such as the FCC in the U.S., Ofcom in the U.K., or Industry Canada), national statistical offices (such as the Office for National Statistics in the U.K. or Statistics Canada), and international agencies (such as OECD, ITU, and the United Nations).¹¹ However, the collected administrative data and statistics first and foremost focus on the access and usage of devices, and do not discriminate between their performances. A simple mobile phone is not distinguished from an internet-enabled smart phone and a dial-up internet connection not from a broadband fiber optic cable. In a world where it is increasingly about better, not more technology, this can be misleading when drawing policy conclusions on basis of these statistics.

For example, the digital divide is defined as the divide between already included into the information society, and those marginalized.¹² Based on the available statistics it is usually accounted for by the numbers of ICT devices per person. Since the numbers of devices per capita in the developed world has already reached

a level of saturation (with several countries reaching a mobile phone penetration of over 100 %), the developing countries will inevitably catch up, the divides closes and the logic conclusion would be that we should not worry about it (which is currently the prevailing attitude: market mechanisms alone would result in informational equality). However, as we have seen, the number of devices and subscriptions is ever less systematically related to informational capacity.¹³

The left-hand side of Figure 2 shows the international digital divide in terms of fixed-line phone and fixed-line Internet subscriptions per inhabitant between the developed member countries of the OECD, and non-OECD developing countries. The divide has gradually been closed over the last decade, from 7 : 1 to 3.5 : 1. The right-hand side shows the same, but this time in terms of the telecommunication capacity in optimally compressed kbps. In the year 2001, the average inhabitant of the developed OECD countries counted with an installed telecommunication capacity of 32 kbps, while the average inhabitant of the developed world merely has access to 3 kbps. A decade later, the member of the developing world had access to almost ten times as much telecommunication capacity (275 kbps), but in the meantime the average inhabitant of the developed world multiplied its access capacities by a factor of one hundred (3,200 kbps). The divide in absolute terms grew from 29 kbps to about 2,900 kbps. Driven by incessant technological progress, the digital age behaves like the Red Queen in Alice's Wonderland, whereas "it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"¹⁴

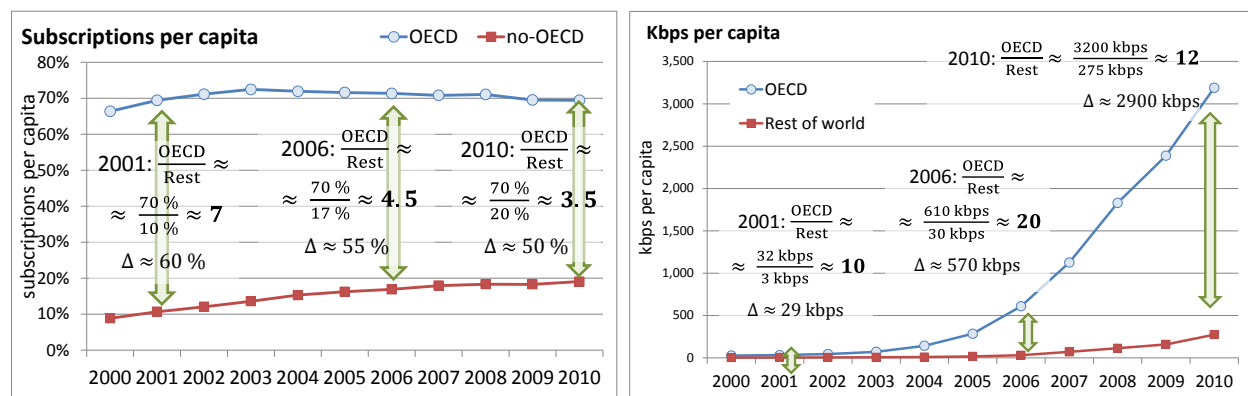


Figure 2. International digital divide in terms of telecom subscriptions per capita (left) and optimally compressed kbps of telecom capacity per capita (right). Taken from⁹

While our technological capacities to handle information have certainly become mind-boggling, compared to the orders of magnitude with which nature processes

information we are still but humble apprentices. In 2007, the DNA in the 60 trillion cells of one single human body stores more information than all of our technological devices together (in both cases information is highly redundant). One hundred human brains can roughly execute as many nerve impulses as our general-purpose computers can execute instructions per second, and the inner circulatory system of only 1,000 people send as many blood cells around per second as human kinds sends bits around. This implies that we live through a time during which we are reaching the extraordinary orders of magnitude with which mother nature processes information in order to sustain intelligent life. What does this mean?

Authors from Hollywood to the scientific community have spilled much ink writing about the imminent technological singularity, a point of greater-than-human intelligence through technological means. While there are certainly profound changes ahead, in the meantime, that is, during the next century or so while this will be unfolding, we can also turn this question around and reasonably ask: if it is true that one human body has an informational capacity that is roughly in the same order-of-magnitude ballpark than all of our technological devices together, why is it that we currently spend US\$ 3.5 trillion per year on our information and communication technology, but less than US\$ 50 on the primary education of a child in many parts of Africa? As a social scientist the inevitable question is: Why do we leave all of this biological information processing capacity unused right now? And what would happen to social evolution if we would finally start to explore the entirety of humankind's innate informational capacity and then combine this full available potential of our uniquely human intelligence with our predictably growing technological capacity?

Martin Hilbert is Provost Fellow at the Annenberg School of Communication at University of Southern California, and Economic Affairs Officer of the United Nations (currently at the U.N. Economic Commission for Latin America and the Caribbean, UN-ECLAC, Chile). He pursues a multidisciplinary approach to understanding and explaining the role of information, communication and knowledge in complex social systems, especially for development (more <http://www.martinhilbert.net>).

¹ The complete collection of the different studies can be access online free of charge through:

<http://www.martinhilbert.net/WorldInfoCapacity.html>

² López, P., & Hilbert, M. (2012). Methodological and Statistical Background on The World's Technological Capacity to Store, Communicate, and Compute Information. *Online document*. For more on the methodology see also endnotes 6 and 7.

³ Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65.

⁴ Hilbert, M. (2011). *That giant sifting sound*. Online video animation, The Economist, <http://ideas.economist.com/video/giant-sifting-sound-0>

⁵ Hilbert, M. (forthcoming). What is the content of the world's technologically mediated information and communication capacity: how much text, image, audio and video?

⁶ Hilbert, M., & López, P. (2012). How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part I: results and scope. *International Journal of Communication*, 6, 956–979.

⁷ Hilbert, M., & López, P. (2012). How to Measure the World's Technological Capacity to Communicate, Store and Compute Information? Part II: measurement unit and conclusions. *International Journal of Communication*, 6, 936–955.

⁸ See for example Gantz, J., et al., (2008). The Diverse and Exploding Digital Universe. IDC (International Data Corporation) sponsored by EMC. Also Bohn, R., & Short, J. (2009). How Much Information? 2009 Report on American Consumers. University of California, San Diego.

⁹ Hilbert, M. (2011). Mapping the dimensions and characteristics of the world's technological communication capacity during the period of digitization. 9th World Telecommunication/ICT Indicators Meeting, *International Telecommunication Union (ITU)*.

¹⁰ Hilbert, M. (forthcoming). Information, communication or technology?

¹¹ Partnership for Measuring ICT for Development, (2008). *The Global Information Society: a Statistical View*. Consisting of 7 United Nations agencies, OECD, World Bank, and EUROSTAT.

ITU (International Telecommunication Union). (2011). *Measuring the Information Society 2011*. ITU-D. OECD (Organisation for Economic Co-operation and Development). (2011). *OECD Guide to Measuring the Information Society 2011*. Paris: OECD Secretary General.

¹² Hilbert, M. (2011). The end justifies the definition: The manifold outlooks on the digital divide and their practical usefulness for policy-making. *Telecommunications Policy*, 35(8), 715–736.

¹³ Hilbert, M. (forthcoming). Global, international and national inequality of the world's technological information capacity between 1986 and 2010

¹⁴ Carroll, Lewis. 1917. *Through the looking-glass, and what Alice found there*. Rand, McNally & Company. p. 39.