

Information Capacity**IJoC**

How to Measure “How Much Information”? Theoretical, Methodological, and Statistical Challenges for the Social Sciences

Introduction

MARTIN HILBERT¹University of Southern California
United Nations ECLAC

The question of “how much information” there is in the world goes back at least to the time when 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” (Aristeas, ca. 200 BC, in Charles, 1913, Section 9). In 1949, one year after his seminal (1948) publication that both created and solved most fundamental problems of information theory, the intellectual father of what is known today as the “information age,” Claude Shannon took a pencil and a piece of notebook paper and estimated the order of magnitude of the largest information stockpile he could think of. He used his newly proposed measure of information (which was at that time, quite unknown) called “the bit,” and estimated the Library of Congress to contain some 10^{14} bits (Gleick, 2011, p. 232). Pressed by the exploding number of information and communication technologies (ICTs) that followed the theories of Shannon and his colleagues during the decades to come, several research projects have taken up this question more systematically since the 1960s. In the eight articles of this Special Section, authors of some of the most extensive of those inventories discuss findings, research priorities, advantages, and limitations, as well as methodological and measurement differences in their approaches. The goal is to provide an open and transparent academic dialogue that deepens the understanding of the nature, assumptions and limitations of these kinds of inventories and to create a solid fundament for potential future exercises of a similar kind.

¹ The author of this introductory article thanks István Dienes, Russell Neuman, Yong Jin Park, Elliot Panek, and Andrew Odlyzko for their detailed comments and suggestions. While these comments surely improved the article, all errors, ambiguities, and biases that remain in the article are the sole responsibility of the author.

Martin Hilbert: mhilbert@usc.edu

Date submitted: 2011–06–13

Copyright © 2012 (Martin Hilbert). Licensed under the Creative Commons Attribution Non-commercial No Derivatives (by-nc-nd). Available at <http://ijoc.org>.

As guest editor of this Special Section, I start by providing some of the main conclusions that I draw from this exercise. While these conclusions are my own personal interpretation, and while I bear sole responsibility for eventual blunders, they are informed by the statements made in the various contributions to this section. The goal of these conclusions is to offer the reader a quick overview about the current state of the art, as well as some of the recurrently mentioned challenges (a much more detailed and balanced description of the challenges will be found within the different articles). I also review the historical context of the most well-known and extensive of these inventories, which will provide the reader with the necessary background for the articles in this Special Section. In the final part of this introductory article, I present the eight studies included in this Special Section.

Some Conclusions of This Special Section

While the contributions are diverse, a general reading allows for several conclusions that are discussed in one way or another in most of the contributions to this Special Section:

1. ***It is not only statistically feasible, but also analytically insightful to quantify the amount of information handled by society.***

The main reason social scientists have been slow with the direct measurement of information stocks and flows is that the data sources for such estimations are not readily available. It is much more convenient to resort to proxies, such as the number of devices, industry size, or dollars spent, which are already harmonized in available databases. Counting bits and bytes requires drawing from a large variety of scattered sources. Notwithstanding, the existing inventories show that the necessary ingredients to create these statistics are available, and that several substantively interesting insights can be gained that cannot be obtained without the direct accounting of information.

2. ***However, many of the available sources are not very solid, and the methodologies are still maturing.***

Despite this undisputed success, the authors in this Special Section also stress that the available data sources and methodological assumptions have many limitations. This should not be surprising, since many of the used statistics have not been created for information inventories, and the applied methodologies are relatively young. The result is far from the institutionalized mechanisms, elaborate manuals, and financial backing many other globally relevant indicators can count on. For now, information inventories are full of trial-and-error concepts, and the result is often a compromise between what is methodologically desirable and what is statistically feasible.

3. ***The research question and its theoretical framework have defined the methodology, including the choice of the indicator.***

While the previous comment applies in general, the heterogeneous contributions of this Special Section leave no doubt that several of the methodological differences between approaches are not rooted in methodological immaturity, but simply in difference in the research focus. Some

studies focus on the amount of information supplied by producers, others on the amount of information demanded by consumers (or both). Some inventories aim at quantifying the hardware capacity of technologies, while others quantify the amount of (optimally compressed) information contained in that hardware. Some studies focus on the installed capacity of information storage versus communication, while others focus on the final consumption of information, regardless of its stored or communicated origin. Some studies aim for counting only original and unique information, while others do not distinguish between duplicate and original. Such differences account for the vast majority of the differences in the resulting numbers. As always, the chosen measurement indicator is defined by the chosen theoretical framework and the particular research question on the researcher's mind.

4. **There is still no consensus on how to define the most fundamental measures for data and information.**

Porat (1977) popularized the definition that "information is data that have been organized and communicated" (p. 2). Bounie and Gille (this Special Section) also apply a broad definition and understand information "in the broadest sense, combining notions of data, intellectual works, media, etc." Bohn and Short (this Special Section) "define data as artificial signals intended to convey meaning, and information as data that is actually received by a person." Hilbert and López (Part II, this Special Section) become more technical and define data as the hardware symbols that physically contain information (which they account for in "binary digits"), and information as the part of this hardware capacity that is optimally compressed and stripped of its redundant parts ("Shannon's entropic bits"). As a result, some studies quantify information in terms of hardware binary digits, optimally compressed bits, number of word equivalents, or number of hours of consumption. Again, each of these units of measurement emphasizes some aspects and silences others.

5. **Information quantity is not equal to information quality or information value, but the second requires the first.**

Many of the authors stress that the quantification of information does not necessarily say anything about the quality or value of this information. At the same time that many lament our lack of understanding of the value of information (or the monetization of that value), the articles in this Special Section make clear that we do not even have a clear understanding of the nature and role of the quantity of information. This is curious, because per definition of "value OF information" or "quality OF information," any quantifiable measure of value and quality will first of all require a quantifiable measure of information: [value of information / amount of information], or [quality / unit of information]. In order to create indicators such as [US\$ / bit], [attention / bit], or [pleasure / bit], one first of all needs to measure the denominator of the ratio: the amount of information. In order to test hypothesis about the value of information, we have to answer the "how much information" question first. Without normalization on the quantity of information, we would helplessly confuse the effects of "more information" with those of "better information." In this sense, the quantification of information enables us to narrow down what we mean by "quality" or "value." Based on this, we will then be able to analyze hypotheses like "the value of video per bit is much lower than the value of text per bit," or "there are

decreasing returns of a certain nature to a bit-flood." Without information inventories, these hypotheses are mere speculations.

6. ***Will it be possible and/or useful to harmonize information accounts?***

A careful reading of the articles in this Special Section allows us to distinguish two future visions. On the one hand, Part I of the contribution of Hilbert and López provides many examples in which the "question on the researcher's mind" clearly defines how to go about measuring information. They conclude that any methodological decision and any choice of metric have been taken in response to a particular research focus (see also conclusion 3), and they suggest that this will continue to be this way for the foreseeable future. On the other hand, the contribution of Bounie and Gille and the article by Dienes suggest that it is desirable to work as directly as possible toward a harmonization of the different methodologies. Bounie and Gille talk about the creation of satellite accounts² for information inventories to complement the national and international statistical data machinery, while Dienes (inspired by the success story of the System of National Accounts, or SNA,³ over recent decades) even suggests a System of National Information Accounts (SNIA) that would harmonize the measurement of stocks and flows of information. History has taught us that it is useful to set up an institutional mechanism to regularly collect important and influential indicators, and harmonized methodologies are certainly be required in order to do so. The Organisation for Economic Co-operation and Development (OECD) has long maintained a Working Party on Indicators for the Information Society (OECD, 2011). While this working party does not measure information directly (in the sense of the inventories included in this Special Section) but mainly works with proxies for information stocks and flows (such as number of devices and respective spending), the sustained work of the OECD shows that there is a broad international interest in and commitment to collecting information indicators. The well-known drawback of institutionalized statistics creation is their inertia, which often leads to the creation of obsolete or meaningless indicators over time and their potential to bias our understanding of a certain issue by blending out alternative ways of looking at it (e.g. see Stiglitz, Sen, & Fitoussi, 2009 for a critique of GDP and related economic indicators). To minimize this risk, it is advisable that methodological choices are very mature and solid before they are fed into the global statistical machinery. Either way, this Special Section provides the first international forum to compare the existing approaches and to work toward a maturation of the applied methodologies, independently of the question if information inventories will eventually be harmonized and institutionalized or not.

² In statistical data gathering, satellite accounts provide a framework that enables attention to be focused on a certain field or aspect of economic and social life in the context of national accounts; common examples are satellite accounts for the environment, tourism, or unpaid household work, which are linked to the central accounts.

³ The System of National Accounts (SNA) is the internationally agreed-upon standard set of recommendations on how to compile measures of economic activity. The SNA describes a coherent, consistent, and integrated set of macroeconomic accounts in the context of internationally agreed-upon concepts, definitions, classifications, and accounting rules.

Introduction: History and Context

The modern-day fascination of social scientists with inventories of social information and communication goes at least back to Machlup's groundbreaking work *The Production and Distribution of Knowledge in the United States* (1962). Following the logic of national accounting in economics, Machlup identified those sectors of the economy that he (quite subjectively) considered to be information- and knowledge-intensive and tracked the size of the respective industries (in US dollars) and occupational force. Following Machlup's lead, Porat (1977) evolved this approach. He famously concluded that the value of the composed labor and capital resources of these "information" sectors made up 25% of U.S. gross domestic product in 1967. This estimate is based on a rather subjective identification of "information capital" and "information workers." He measures the economic value of the related "information activity [which] includes all the resources consumed in producing, processing, and distributing information goods and services" (p. 2). As information capital he loosely identified a "wide variety of information capital resources [which] are used to deliver the informational requirements of one firm: typewriters, calculators, copiers, terminals, computers, telephones and switchboards . . . microwave antennae, satellite dishes and facsimile machines" (pp. 2-3). Despite all coarse-graining and methodological arbitrariness of this approach to represent the role of information in an economy, Machlup and Porat's work constitute important milestones with regard to evaluating the economic dimensions of information in a society.

Another approach does not focus on the economic output of information sectors, but on the amount of information itself, independent of its assigned value. Here the first efforts date back to the 1970s and early 1980s and were championed by Japan's Ministry of Posts and Telecommunications (MPT) (Ito, 1981). In 1975, a so-called Information Flow Census was carried out by Japan's MPT (for a good summary in English, see Duff, 2000). The census aimed at obtaining empirical evidence of the volume and vehicles of information in circulation in Japanese society. The ministry created statistics for a broad collection of electronic technologies (including telephone, telegraph, data communication, facsimile, radio, TV, tape, and record), as well as non-electronic technologies (including postal mail, newspaper, and book), and even accounted for direct human communication in the classroom and conversations outside the home. Initially the authors chose binary digits as the unit of measurement, that is, the number of 1s and 0s involved when operating those technologies. However, they felt that the results did not sufficiently recognize the contribution of text, in relation to data-intensive images and voice. The transmission of images requires more binary digits than does the transmission of plain text, especially when little compression is used, as was the case in the 1980s. As a result, the authors decided to introduce the measure of "amounts of words" as the unifying unit. This was effectively implemented by the definition of conversion rates between informational content and the corresponding number of words, which included various more-or-less disputable assumptions. Based on the best sources available at the time, it was assumed that a minute of speech over radio or a telephone line was equal to 120 words, a picture on a fax machine was equal to 80 words per page, and TV provided 1,320 words per minute, which also applies to cinema and face-to-face conversations in school education. The census distinguished between supply of information (the amount of information sent out) and consumption of information (the amount of information read or listened to).

Some interesting conclusions can be drawn from this pioneering effort (Ito, 1981). The most convincing results concern the analysis of trends over time, since growth rates can reveal relative tendencies quite independently of the chosen unit of measurement, as long as the chosen unit of measurement is applied consistently. It was shown that electronic media and two-way-personal-communication media had become much more price-effective, while non-electronic mass media had stagnated or had even become more expensive (Duff, 2000). It was also shown that the amount of information supply was increasing much faster than the amount of information consumption, providing the first empirical evidence of what nowadays is commonly recognized as "information overload."

In the early 1980s, Ithiel de Sola Pool and collaborators (1984) decided to broadly adopt the Japanese methodology, which led to a Japan-USA comparative study. Pool ignored all imagery and music and counted only the actual words transmitted, as well as their price. Although this reduces some of the most disputable assumptions, it leaves out a major part of the information revolution. The parts of the results that correspond to the Japanese exercise are very similar and reconfirm the previous findings. The message that point-to-point communication through electronic media was becoming the dominant form of communication even made it into the prestigious journal *Science* (Pool, 1983). However, as was its predecessor, this study was criticized for its choice of indicator, which focuses on text. Ironically, the choice of indicator was contradictory to the general trend in technological information processing: "If one thing is clear about post-industrialism and the information society it is surely that the 'hegemony' of text is eroded" (Duff, 2000, p. 87).

In the meantime, a noteworthy effort was under way in Hungary. In 1981, the Hungarian Central Statistical Office started a research project to account for the country's information industries, including libraries, education, culture, mass media, health services, and government agencies. Thanks to the centralized socialist system of official statistics, the office had access to a large variety of sources. In 1986, the office announced a new branch of official statistics, called "information statistics," whose ultimate purpose was to install information activities into the system of "national accounts." It published definitions of concepts like information good, service, activity, industry, and nomenclatures and issued the first publications, which described information economy as a whole and by industries (Dienes, 1986). Information activities and their resources in Hungary were measured in value and volume (bit) terms, and by 1990 the first bit-term balances of information were compiled, with output, consumption, use, exports, imports (with foreign externalities), and stocks of information products all together and by kinds. In the early 1990s, István Dienes, chief scientist of the Hungarian Central Statistical Office, compiled a manual for a standard system of national information accounts (SNIA) (Dienes, 1993). He also compiled a draft of a bit-term sectorial balance for the U.S. (Dienes, 1994b). After the fall of the Soviet bloc in 1989, the effort was reduced, but it still persists in Hungary and consequently is the longest lasting in this field (Dienes, 2010).

In the Western world, it took almost two decades before this approach was taken up again. In 1997, Michael Lesk, a professor of library and information science, posed the intriguing question "How Much Information Is There in the World?" He provided a short, 10-page outline on how to go about estimating it. His focus was notably set on storage much more than on communication, and his unit of measurement was the hardware capacity of available storage devices (binary digits).

A group of researchers at the University of California, Berkeley, at what is now the School of Information, took up the measurement challenge for the years 2000 and 2003. Peter Lyman and Hal Varian (2000, 2003) led two groundbreaking studies with the characteristic title "How much information?" The studies aimed at obtaining two year-bound inventories of the quantity of information that existed worldwide. In the words of the authors (2000, p. 3):

We have identified production of content by media type, translated the volume of original content into a common standard (terabytes), determined how much storage each type takes under certain assumptions about compression, attempted to adjust for duplication of content, and added up to get total estimates.

As a result, the studies show estimates for the amount of uniquely created information stored on paper, film, magnetic media, and optical devices, and the amount of uniquely created bits flowing through broadcasting, telephony, and the Internet, measured in the number of binary digits (0s and 1s) that represent this information. The remarkable results showed that 92% of new information was stored on magnetic media, primarily hard disks; that electronic channels (telephone, radio, TV, and the Internet) contained 3.5 times more unique information than what was recorded in storage media; and that the United States produced about 40% of the world's newly created information that was eventually stored in some kind of device.

These estimates have been refined for the case of the European Union by Bounie (2003). The study reconfirmed many of the findings of Lyman, Varian et al. (2000, 2003) for the estimation of the flow and the stock of original content, as well as for the flow and the stock of copies. In addition, Bounie estimated the monetary values of those flows and stocks and found that the European Union captured 36% to the global turnover of 349 billion euros, and the United States contributed with 58% of this global total (mainly TV broadcasting), despite the fact that the flow of original content was to the advantage of the European Union. This raises the interesting question of how to place monetary values on the flow of information.

The thought-provoking insights of these studies awakened the interest of the information and communication technology industry. The storage company EMC commissioned the private-sector research firm IDC to track the size of the information flowing through the "digital universe" for the years 2007 and 2008 (Gantz et al., 2008). The researchers estimated that in 2007 "all the empty or usable space on hard drives, tapes, CDs, DVDs, and memory (volatile and nonvolatile) in the market equaled 264 exabytes" (p. 4).⁴

⁴ The estimate by Hilbert and López (2011) of the hardware capacity of merely the digital part of global information storage amounts to 363 exabytes of binary digits, significantly larger than the IDC estimate. The Hilbert and López inventory is more comprehensive; however, it is unlikely that this explains the entire difference. Unfortunately, given the proprietary nature of IDC's work, the adopted working assumptions of the IDC study are often not explained in detail and the relevant statistics cannot be

Table 1. Conventions and Prefixes

Conventionally, bits are abbreviated with a small "b" (such as in kilobits per second: kbps) and bytes (equal to 8 bits) with a capital "B" (such as in megabyte: MB).		
Kilo	$10^3 = \text{thousand}$	x,000
Mega	$10^6 = \text{million}$	x,000,000
Giga	$10^9 = \text{billion}$	x,000,000,000,
Terra	$10^{12} = \text{trillion}$	x,000,000,000,000
Peta	$10^{15} = \text{quadrillion}$	x,000,000,000,000,000
Exa	$10^{18} = \text{quintillion}$	x,000,000,000,000,000,000
Zetta	$10^{21} = \text{sextillion}$	x,000,000,000,000,000,000,000
Yotta	$10^{24} = \text{septillion}$	x,000,000,000,000,000,000,000,000

In parallel, a unique, longtime effort by the Minnesota Internet Traffic Studies project (Odlyzko, 2009) has assessed the global flow of data through the Internet backbone measured in bandwidth of binary digits. The global IP traffic has also been estimated by Cisco Systems (2008). Choosing a similar focus on telecommunications, but measuring traffic in terms of minutes, several social-network analyses looked at the international flow of Internet traffic (Barnett, Chon, & Rosen, 2001) or fixed-line telephony traffic (Monge & Matei, 2004; Seungyoon, Monge, Bar, & Matei, 2007).

Neuman, Park, and Panek (see this Special Section) returned to Pool's original methodology and concentrated on the question of information overload. To better address this specific question, they reported the final metric in minutes, not in the number of words, as Pool had done. The methodology measured how much information was supplied (for example, how many TV channels) and compared it with how much information was consumed (for example, the typical TV displays only one TV channel at a time), which provides a rough indicator for the variety of information content (minutes offered versus minutes consumed). The results show that the ratio between supply and demand grew from 82 : 1 minutes in 1960 (which is supplied by 3.4 television stations, 8.2 radio stations, 1.1 newspapers, 1.5 recently purchased books, and 3.6 magazines per household), to 884 : 1 minutes in 2005. In other words, per minute of media consumption available, one can choose from 884 minutes provided. They conclude that this "is *not* a human-scale cognitive challenge; it is one in which humans will inevitably turn to the increasingly intelligent digital technologies that created the abundance in the first place for help in sorting it out" (Neuman, Park, & Panek, this Special Section).

In 2008, another university-industry consortium started to research the "how much information" question at the University of California, San Diego. The first study by this group was on information

checked, since they are mostly taken from inaccessible company sources. This makes it difficult to evaluate the validity of or to replicate the IDC results.

consumption in U.S. households in 2008 (Bohn & Short, 2009). This effort focused explicitly on information consumption, without distinguishing between information retrieval from a storage device and information delivery over a communication network. Both methods eventually led to information consumption. Three units were used to measure consumption: hardware binary digits, words, and hours. Given the focus on consumption, the results are very sensitive to the relevant media-consumption studies, which aim at estimating how many minutes people interact with a media device. Based on this approach, the U.S. study found that TV, computer games, and movies represented 99.2% of the total number of bits "consumed." In 2011, the group published an estimate of how much information was processed by the world's enterprise servers (Short, Bohn, & Baru, 2011). It took into account that "a single chunk of information, such as an e-mail message, may flow through multiple servers and be counted multiple times" (p. 7) and focused on the effective processing of information, not the installed capacity. The group found that two-thirds of the world's total of 9.57 zettabytes was processed by low-end, entry-level servers costing US\$25,000 or less.

Hilbert and López (2011) also estimated the world's computational capacity, but in this case in millions of instructions per second (MIPS), another measure of computational hardware capacity. They found that the hardware capacity of humanly guided general-purpose computation grew at an impressive compound annual growth rate of 61% between 1986 and 2007, and that embedded applications-specific computation grew even faster, at 86%. In the same exercise, Hilbert and López also took inventory of the world's technological capacity to store and communicate information in bits between 1986 and 2007, providing consistent long-term time series for more than 60 categories of analog and digital technologies. To be able to do this, they had to harmonize the amount of storage and communication capacities of the available hardware with compression rates. The resulting logic is similar to what economists are accustomed to when normalizing for inflation: The creation of meaningful time series for analog and digital technologies requires the normalizing of data with different levels of redundancy on one chosen level of compression. Hilbert and López treat all information as if it were compressed with the most efficient compression algorithm available in 2007, a measure they call "optimally compressed bits." The maximum level of compression has a special status, since Shannon (1948) has proven that the uttermost compression of information approaches the entropy of the source. They found that the world's storage capacity grew at a compound annual growth rate of 25% per year between 1986 and 2007, and the world's telecommunication capacity at 30% per year.

Summary and Differences in Focus

This leaves us with several broad methodological choices, with several different units of measurement, leading to complementary results. The main differences in focus of existing studies include:

- 1. Differences in distinction of information activity:**

The different technologies can be grouped in different collections, most commonly classified by a specific activity done with the information. Lyman, Varian, et al. (2000, 2003) differentiate between "stocks" and "flows." Hilbert and López (2011) distinguish among information transmitted in space (communication), through time (storage), and transformation of information

(computation). Bohn and Short (2009); Short, Bohn, and Baru (2011); and Neuman, Park, and Panek (this Special Section) focus on information consumption, regardless of whether it originates from storage or communication devices. Other classifications have been proposed by Dienes (1993), such as intermediate consumption, exports, imports, accumulation, human information services, and knowledge embodied in brains, but have not been collected.

2. **Differences in the main unit of measurement:**

Machlup (1962) and Porat (1977) accounted for output in monetary value. The Japanese exercises (Ito, 1981) and Pool (1983) measure information in the equivalent of words. Lyman, Varian, et al. (2000, 2003) account for unique information at different levels of compression, while Lesk (1997) and Gantz et al. (2008) account for the installed hardware capacity. Neuman, Park, and Panek (this Special Section) measure minutes, and Bohn and Short (2009) present three numbers: hardware capacity, word equivalents, and minutes. Hilbert and López (2011) measure the capacity of handling optimally compressed information.

3. **Differences in analyzed sectors:**

Some of the inventories are global and do not fine-grain to distinguish between different sectors, while others do. Machlup (1962) and Porat (1977) worked with hand-selected industries. Bohn and Short (2009) focused their analysis of information consumption at households, and Short, Bohn, and Baru (2011) focused their analysis of computer servers on enterprises. Of course, the overall pie could also be cut differently, and one could focus on specific sectors of the economy, or government, or different social groups.

There are more differences, and throughout this Special Section, authors will compare and contrast the different approaches and methodologies.

Content of This Special Section

This Special Section includes 8 contributions from 7 research teams, consisting of 12 authors. We start with five articles that present the results and methodological decisions of four different "How much information" inventories. The articles are presented in chronological order according to the age of the applied methodology in each inventory.

The first contribution comes from W. Russell Neuman, Yong Jin Park, and Elliot Panek: "Tracking the Flow of Information Into the Home: An Empirical Assessment of the Digital Revolution in the U.S. from 1960 to 2005." It goes back to the original methodology of Pool (1983), and in fact the senior author of the article was a colleague of Pool's at the time of his original inventories and published a related piece on the original data with Pool (Neuman & Pool, 1986).

David Bounie and Laurent Gille take up the methodology of the Berkeley inventories (Lyman, Varian, et al., 2000, 2003) and present their results in "International Production and Dissemination of

Information: Results, Methodological Issues, and Statistical Perspectives.” In addition, they discuss potential future work.

The article by Roger Bohn and James Short presents part of the comprehensive inventory undertaken by their group and focuses on “Measuring Consumer Information Consumption” in the United States, in words, hours, and bytes.

The fourth and fifth articles are parts I and II of the contribution by Martin Hilbert and Priscila López titled “How to Measure the World’s Technological Capacity to Communicate, Store, and Compute Information?” Part I is titled “Results and Scope” and focuses on the outcome of their inventory and on the main methodological decision that had to be made to create these results. It explains alternative approaches that could have been taken. Part II is titled “Measurement Unit and Conclusions.” It focuses mainly on the authors’ metric of choice (optimally compressed bits) and discusses what this indicator can and cannot explain. The articles are supported by an almost 300-page-long supporting appendix that can be accessed at <http://www.martinhilbert.net/WorldInfoCapacity.html>.

The next two articles do not directly present the results of inventories, but rather ask questions about the purpose, validity, and potential changes in focus of the existing approaches. Andrew Odlyzko’s contribution discusses “The Volume and Value of Information,” while Michael Lesk focuses on the issue of “One in a Million: Information vs. Attention.”

The last article in this Special Section is by István Dienes, who authored various information inventories, mainly in Hungary. It is called “A Meta Study of 26 ‘How Much Information’ Studies: Sine Qua Nons and Solutions” and provides a more detailed comparison of several of the exercises.

References

- Aristeas. (ca. 200 B.C.). The letter of Aristeas to Philocrates. (R.H. Charles, Trans., 1913.) Retrieved from <http://www.attalus.org/translate/aristeas1.html>
- Barnett, G., Chon, B.-S., & Rosen, D. (2001). The structure of the Internet flows in cyberspace. *Networks and Communication Studies NETCOM*, 15(1-2), 61-80.
- Bohn, R. E., & Short, J. E. (2009). *How much information? 2009: Report on American consumers*. Global Information Industry Center at the Graduate School of International Relations and Pacific Studies, 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-paritech.fr/bounie/documents/Recherche/Annex.pdf>
- Cisco Systems. (2008). *Global IP traffic forecast and methodology, 2006-2011* (white paper). Retrieved from http://www.hbtf.org/files/cisco_IPforecast.pdf
- Dienes, I. (1986). Magnitudes of the knowledge stocks and information flows in the Hungarian economy. (In Hungarian.) In *Tanulmányok az információgazdaságról KSH-OMIKK*, 1986, pp. 89-101. <http://infostat.hu/publikaciok/86-nagysr.pdf>
- Dienes, I. (1993). Towards a system of national information accounts. Proceedings of the 21st Telecommunications Policy Research Conference, Solomons, Maryland, October 3, 1993. Retrieved from <http://infostat.hu/publikaciok/93-solomonsprez.pdf>
- Dienes, I. (1994a). *National accounting of information* (reference manual of SNIA, Version 1.1). Retrieved from <http://www.infostat.hu/publikaciok/94-ssniav.pdf>
- Dienes, I. (1994b). *Accounting the information flows and knowledge stocks in the U.S.: Preliminary results*. Presentation at the University of California, Berkeley. Retrieved from <http://infostat.hu/publikaciok/94-berkeleyreport.pdf>
- Dienes, I. (2010). Twenty figures illustrating the information household of Hungary between 1945 and 2008. (In Hungarian.) Retrieved from http://infostat.hu/publikaciok/10_infhazt.pdf
- Duff, A. S. (2000). *Information society studies*. London: Psychology Press.
- Gantz, J. F., 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>
- Gleick, J. (2011). *The information: A history, a theory, a food*. New York: Pantheon.
- 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

- 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.
- Lesk, M. (1997). How much information is there in the world? Retrieved from <http://www.lesk.com/mlesk/ksg97/ksg.html>
- Lyman, P., Varian, H.R., Dunn, J., Strygin, A., & Swearingen, K. (2000). *How much information? 2000*. UC Berkeley. Retrieved from <http://www2.sims.berkeley.edu/research/projects/how-much-info>
- Lyman, P., Varian, H. R., Swearingen, K., Charles, P., Good, N., Jordan, L., & Pal, J. (2003). *How much information? 2003*. UC Berkeley. Retrieved from <http://www2.sims.berkeley.edu/research/projects/how-much-info-2003>
- Machlup, F. (1962). *The production and distribution of knowledge in the United States*. Princeton, NJ: Princeton University Press.
- Monge, P., & Matei, S. A. (2004). The role of the global telecommunications network in bridging economic and political divides, 1989 to 1999. *Journal of Communication*, 54(3), 511–531. doi:10.1111/j.1460-2466.2004.tb02642.x
- Neuman, W. R., & Pool, I. S. (1986). The flow of communications into the home. In S. J. Ball-Rokeach & M. G. Cantor (Eds.), *Media, audience, and social structure* (pp.71–86). Beverly Hills, CA: SAGE Publications.
- Odlyzko, A. (2009). *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
- Pool, I. de S., Inose, H., Takasaki, N., & Hurwitz, R. (1984). *Communication flows: A census in the United States and Japan*. Amsterdam: North-Holland and University of Tokyo Press.
- Porat, M. U. (1977). *The information economy: Definition and measurement*. Washington, DC: National Science Foundation, Superintendent of Documents, U.S. Government Printing Office. (Stock No. 003-000-00512-7). Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED142205>
- Seungyoon, L., Monge, P., Bar, F., & Matei, S.A. (2007). The emergence of clusters in the global telecommunications network. *Journal of Communication*, 57(3), 415–434.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423, 623–656. doi:10.1145/584091.584093
- Short, J. E., Bohn, R. E., & Baru, C. (2011). *How much information? 2010: Report on enterprise server information*. Global Information Industry Center at the Graduate School of International Relations and Pacific Studies, UC San Diego. Retrieved from http://hmi.ucsd.edu/howmuchinfo_research_report_consum_2010.php

Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). *The measurement of economic performance and social progress revisited: Reflections and overview*. Commission on the Measurement of Economic Performance and Social Progress. Retrieved from <http://www.stiglitz-sen-fitoussi.fr/en/documents.htm>