hilbert@UCDavis.edu

Quantifying the Data Deluge and the Data Drought

Background note for the World Development Report 2016. World Bank, Washington, DC

April 2015

Martin Hilbert
University of California, Davis (www.martinhilbert.net)

This report contains a 2014 update of the 1986-2007 main graphs presented in

> Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. https://doi.org/10.1126/science.1200970

and was used as input in Spotlight 5 of the 2016 World Bank Report (title: The data revolution)

Table of Contents

hilbert@UCDavis.edu

ABSTRACT:

The world's big data capacity can be understood in terms of the world's storage capacity, and the telecommunication capacity to access this storage ('the cloud').[1] In order to do this we mainly follow the methodology of what has become the standard reference in estimating the world's technological information capacity: Hilbert and López (2011). This methodology has been adopted by others, including by the 2012 ITU report on "Measuring the Information Society" (Chapter 5; ITU, 2012). It shows that the world's technological capacity to store information has increased with a compound annual growth rate of 31 % during the three decades between 1986 and 2014 (from 2.6 exabytes to 4.6 zettabytes), while the world's installed telecommunication capacity has grown with a compound annual growth rate of 35 % during the same period, from 7.5 petabites to 25 exabits)[2].

Methods

The total total capacity is calculated as the sum of the product of technological devices and their information performance. Technological performance is measured in the installed binary hardware digits, which is then normalized on compression rates. The hardware performance is estimated as "installed capacity", which implies that it is assumed that all technological capacities are used to their maximum. For storage this evaluates the maximum available storage space ("as if all storage were filled"), for telecommunication this describes the "end-user bandwidth potential" ("if all end-users would use their full bandwidth"[3]). The normalization on software compression rates is important for the creation of meaningful time series, as compression algorithms have enable to send more information through the same hardware infrastructure over recent decades (Hilbert, 2014a). We normalize on "optimally compressed bits" (as if all content were compressed with the best compression algorithms possible in 2014) (Hilbert and López, 2012b). It would also be possible to normalize on a different standard (e.g. the most used compression algorithms in 2014), but the optimal level of compression has a deeper information theoretic conceptualization as it approaches the entropy of the source (in Shannon's (1948)). For the estimation of compression rates of different content, justifiable estimates are elaborated for 7-years intervals (1986, 1993, 2000, 2007, 2014). For more see (Hilbert, 2015a; López and Hilbert, 2012).

---

[1] Actually, according to Hilbert and Lopez (2011) also the installed capacity to compute information, but quantifying computational capacity is rather tricky (given lack of agreement on useful metrics) and unfortunately no updated numbers exist.

[2] Be aware that –in agreement with common practice—storage capacities are represented in bytes, while telecommunication capacities are accounted for in bits (one byte is equal to 8 bits).

[3] This is merely a "potential", because in reality, negative network externalities create a trade-off in bandwidth among users. For example, estimating that the average broadband connection is 10 Mbps in a given country, does not mean that all users could use this average bandwidth at the same second. The network would collapse.

The world's technological capacity to store information

The estimations for the period 1986 – 2007 follow Hilbert and López (2011). The update for 2007 – 2014 follows a mix of estimates. Several related large scale exercises have been undertaken to tackle the so-called "how much information" question. Hilbert (2015b) reviews the methodologies of the eight most prominent ones in a comparative manner. A special issue of the International Journal of Communication (Vol. 6, 2012) goes more into detail about the differences (see the introduction, Hilbert, 2012, and the eight contributions therein). Considering these methodological differences and more current updates (i.e. Gantz and Reinsel, 2012; Turner et al., 2014), the following estimates were elaborated.

Figure 1 shows that the world's technological capacity to store information has almost reached 5 zettabytes in 2014.[4] This results in a compound annual growth rate of some 30 %. This is about five times faster than the world economy grew during the same period. The digitalization of the world's information stockpile happened in what is a historic blink of an eye: in 1986, less than 1 % of the world's mediated information was still stored in digital format. By 2014, less than 0.5 % is stored in analog media. Some analog storage media are still growing strongly today. For example, it is well known that the long-promised "paperless office" has still not arrived. The usage of paper still grows with some 15 % per year (some 2.5 times faster than the economy), but digital storage is growing at twice that speed. The nature of this exponential growth trend leads to the fact that until not too long ago (until the year 2002) the world still stored more information in analog than in digital format. The year 2002 is often cited as the "beginning of the digital age".

It is useful to put these mind-boggling numbers into context. If we would store the 4.6 optimally compressed zettabytes of 2014 in 730 MB CD-ROM discs (of 1.2 mm thickness) we could build about 20 stacks of discs from the earth to the moon. If we would store the information equivalent in alphanumeric symbols in double printed books of 125 pages, all the world's landmasses could have been covered with one layer of double printed book paper back in 1986. By 1993 it would have grown to 6 pages, and to 20 pages in the year 2000. By 2007 it would be one layers of books that covers every square centimeter of the world's land masses, two layers by 2010/2011, and some 14 layers by 2014 (letting us literally stand "knee-deep in information")[5].

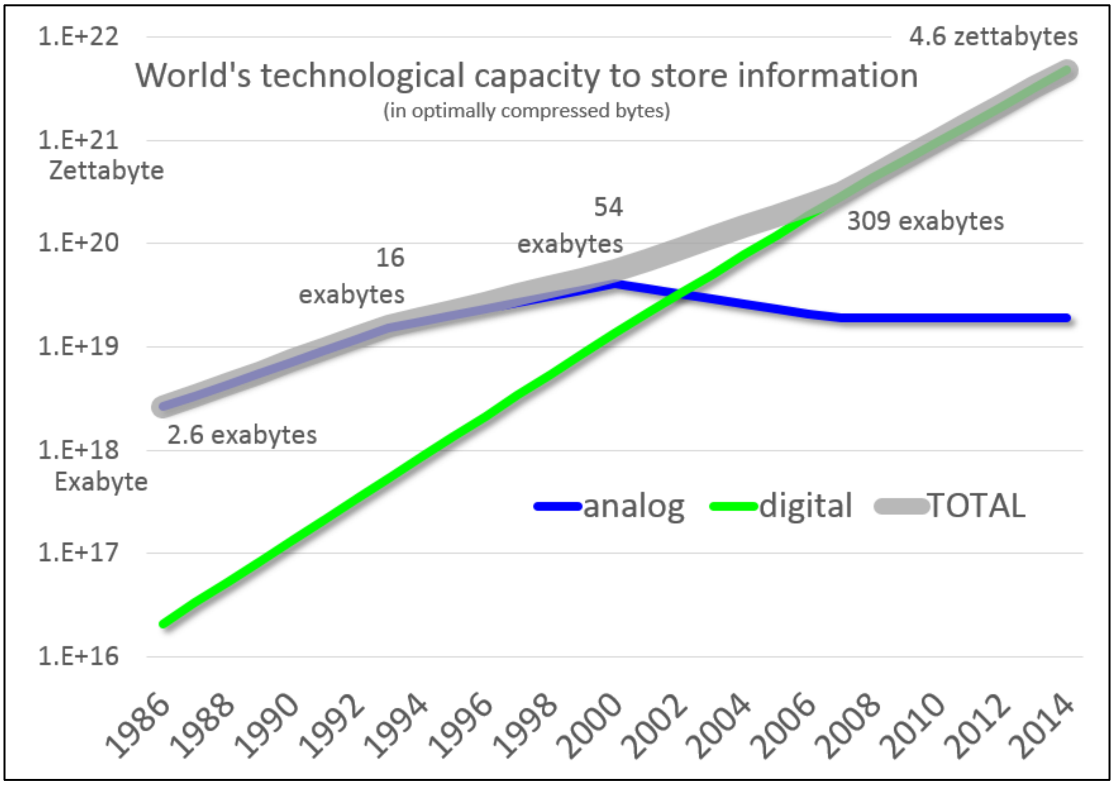
An increasing share of this total storage capacity (some 20 – 40 %) is centralized in "the cloud", with extremely rapidly falling prices for cloud services (Hilbert, 2014b; Turner et al., 2014, Bort, 2014). The gatekeeper to obtain access to this centralized capacity is telecommunication access. Telecommunication channels have the capacity to convert even highly concentrated cloud services into highly democratized digital capacities through decentralized telecommunication networks. Installed telecom bandwidth potential is the necessary condition for connectivity to the Big Data cloud.

---

[4] I's often useful for readers to include a small box with introduction to decimal binary prefixes: http://en.wikipedia.org/wiki/Binary_prefix , i.e. kilo, mega, giga, tera, peta, exa, zetta => each adding '000.

[5] This lends itself to nice info-visualization, much like: https://www.youtube.com/watch?v=iIKPjOuwqHo

Figure 1: World's technological capacity to store information 1986 – 2014 (log on y-axis).



Source: based on the methodology of Hilbert & Lopez, 2011, with own estimates for 2007-2014.

The world's technological capacity to telecommunicate information

In contrary to storage, the installed telecommunication bandwidth potential is based on a non-tradable infrastructure, which makes it straightforward to assign it geographic ownership. We can now look at the evolution of the installed bandwidth potential in optimally compressed kbps among different world regions. For this we include all fixed and mobile telephony and all fixed and mobile internet. The subscriptions data stem mainly from ITU (2014), with completions from other sources, while the estimation of bandwidth potential is based on a diverse collection of sources (see Hilbert, 2015a). One of the main sources for internet bandwidth is NetIndex (Ookla, 2015), which has gathered the results of end-user-initiated bandwidth velocity tests per country per day over recent years (e.g., an average 180,000 test per day already in 2010 through Speedtest.net and Pingtest.net). It is seen as “the best of the currently available data sources for assessing the speed of ISP's broadband access service” (Bauer, Clark,&Lehr, 2010, p. 3).

Figure 2a looks at the total telecommunication capacity in optimally compressed kbps in terms of global income groups (following the classification of the World Bank of 2015). The last three decades show a gradual loss of dominance of global information capacities for today's high income countries. High income countries dominated 86 % of the globally installed bandwidth potential, but merely 66 % in 2013. It is interesting to compare this presentation with the more common method to assess the advancement approximation in terms of the number of telecommunication subscriptions (Figure 2b). Both dynamics are quite different, which stems from the simple fact that not all subscriptions are equal in their communicational performance. This intuitive difference is the main reason why the statistical accounting of subscriptions is an obsolete and very often misleading indicator. This holds especially true in an age of Big Data, where the focus of development is set on informational bits, not on the number of technological devices (for the complete argument, see Hilbert 2014b).

Comparing these results with the global shares of Gross National Income (GNI) and population (Figure 2c and 2d), it becomes clear that the diffusion dynamic of the number of subscriptions follows existing patterns in population distribution. Especially the diffusion of mobile phones during recent decades has contributed to the fact that both distributions align. The number of subscriptions reaches a saturation limit at about 2 to 2.5 subscriptions per capita worldwide (see also Hilbert 2014b), and therefore leads to a natural closure of the divide over time. On the contrary, communication capacity in kbps (and therefore access to the global Big Data infrastructure) follows the signature of economic capacities. After only a few decades, both processes align impressively well. This shows that the digital divide in terms of data capacity is far from being closed, but is rather becoming a structural characteristic of modern societies, which is as persistent as the existing income divide (Hilbert 2014b).

Figure 2: International income groups: (a) telecommunication Capacity in optimally compressed kbps; (b) telecommunication subscriptions

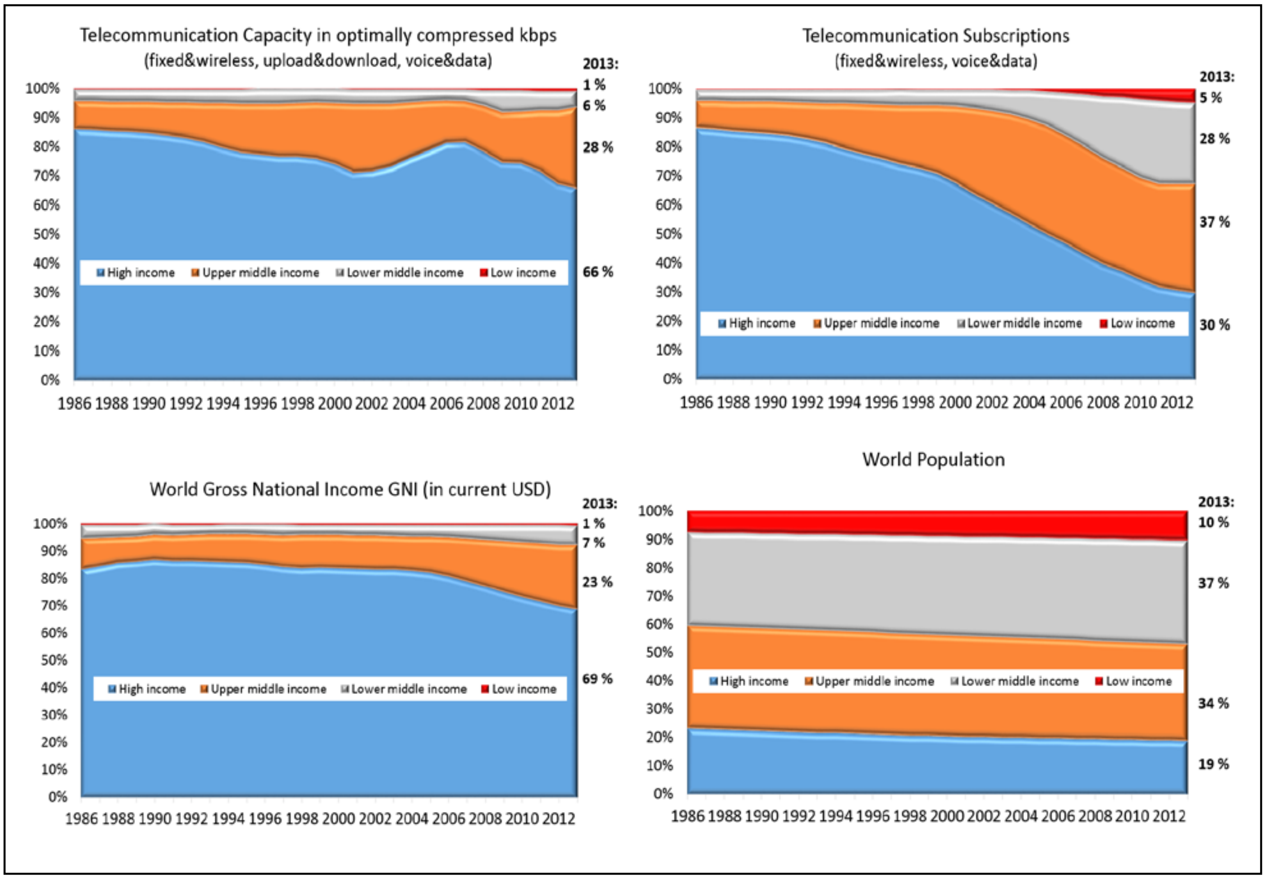
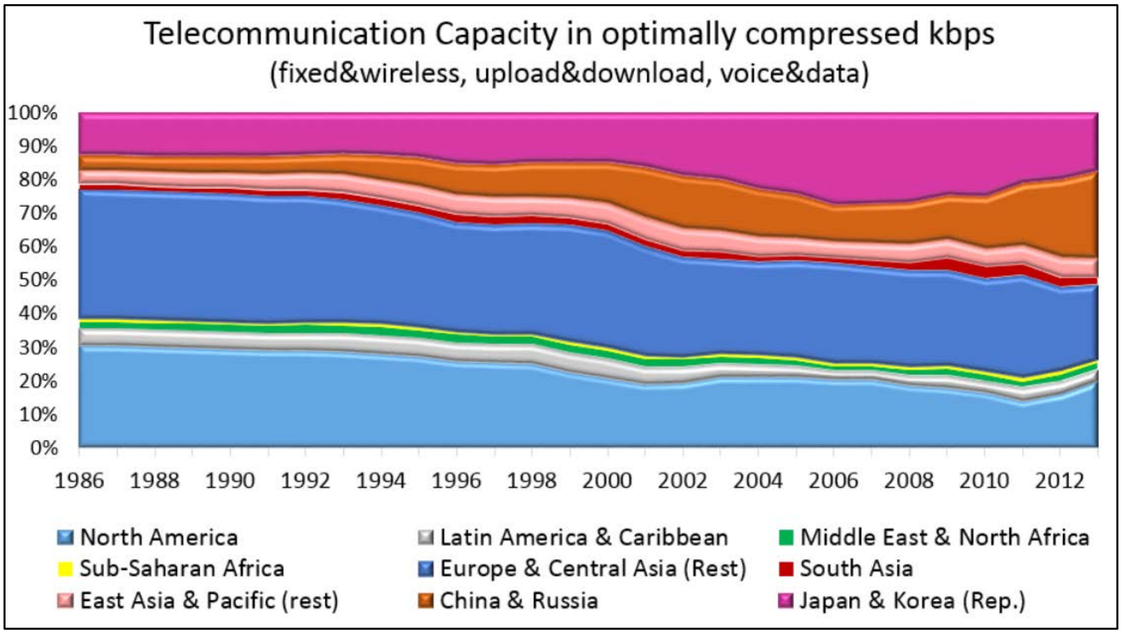


Figure 2a also reveals that the evolution of communication capacities in kbps is not a monotone process. Increasing and decreasing shares between high income and upper middle income countries suggest that the evolution of bandwidth is characterized by a complex non-linear interplay of public policy, private investments, and technological progress. Some countries in this income range seem to (at least temporarily) do much better than their economic capacity would suggest. This is a typical signature of effective public policy.

Figure 3 shows the same global capacity in optimally compressed kbps per geographic regions (following the World Bank classification of 2015). Asia has notably increased its global share at the expense of North America and Europe, with a share of less than a quarter of the global capacity in 1986 (23 %) and a global majority of 51 % in 2013 (red shaded areas in Figure 3). Figure 3 reveals that the main driver of this expansion during the early 2000s were Japan and South Korea, both of which famously pursued a very aggressive public sector policy agenda in the expansion of fiber optic infrastructure in the early 2000s (e.g. Kantei, 2001; Rhee and Kim, 2004). The more recent period since 2010 is characterized by the expansion of bandwidth in both China and Russia. Notably, most recent broadband policy efforts in the U.S. (FCC, 2010) seems to show some first detectable effects on a macro-level, as North America has started to return its tendency of a shrinking global share during recent years.

Figure 3: Telecommunication Capacity in optimally compressed kbps per world region



Expressed in installed kbps per capita (per inhabitant) we can obtain a clearer picture about the increasing and decreasing nature of the evolving digital divide in terms of bandwidth capacity. First in foremost, Figure 4a shows that the divide continuously increases in absolute terms. In 2003, the average inhabitant of high income countries had access to an average of 100 kbps of installed bandwidth potential, while the average inhabitant of the rest of the world had access to merely 9 kbps. In absolute terms, this results in a difference of some 90 kbps. As shown in Figure 4a, this divide increased with an order of magnitude every 5 years, reaching almost 900 kbps in 2007, and over 10,000 kbps by 2013. This increasing divide in absolute terms is important to notice in the context of a Big Data world, in which the amount of data is becoming a crucial ingredient for growth.

In relative terms, this results in an increasing and decreasing evolution of the divide over time. Figure 4b contrasts this tendency with the monotonically decreasing tendency of the digital divide in terms of telecommunication subscriptions. It shows that the divide in terms of data capacities is much more susceptible to both technological change and technology interventions. The decreasing divide during the period until 2000 is explained by the global diffusion of narrowband internet and 2G telephony. The increasing nature of the divide between 2001 and 2008 is due to the global introduction of broadband for fixed and mobile solutions. The most recent decreasing nature of the divide is evidence of the global diffusion of broadband. The digital divide in terms of data capacities is a continuously moving target, which opens up with each new innovation that is introduced into the market (see also Hilbert, 2014b).

Figure 4: (a) Telecommunication capacity per capita in optimally compressed kbps: high income groups (World Bank classification) versus Rest of world. (b) ratio of telecommunication capacity per capita in high income countries versus rest of world, and of subscriptions per capita

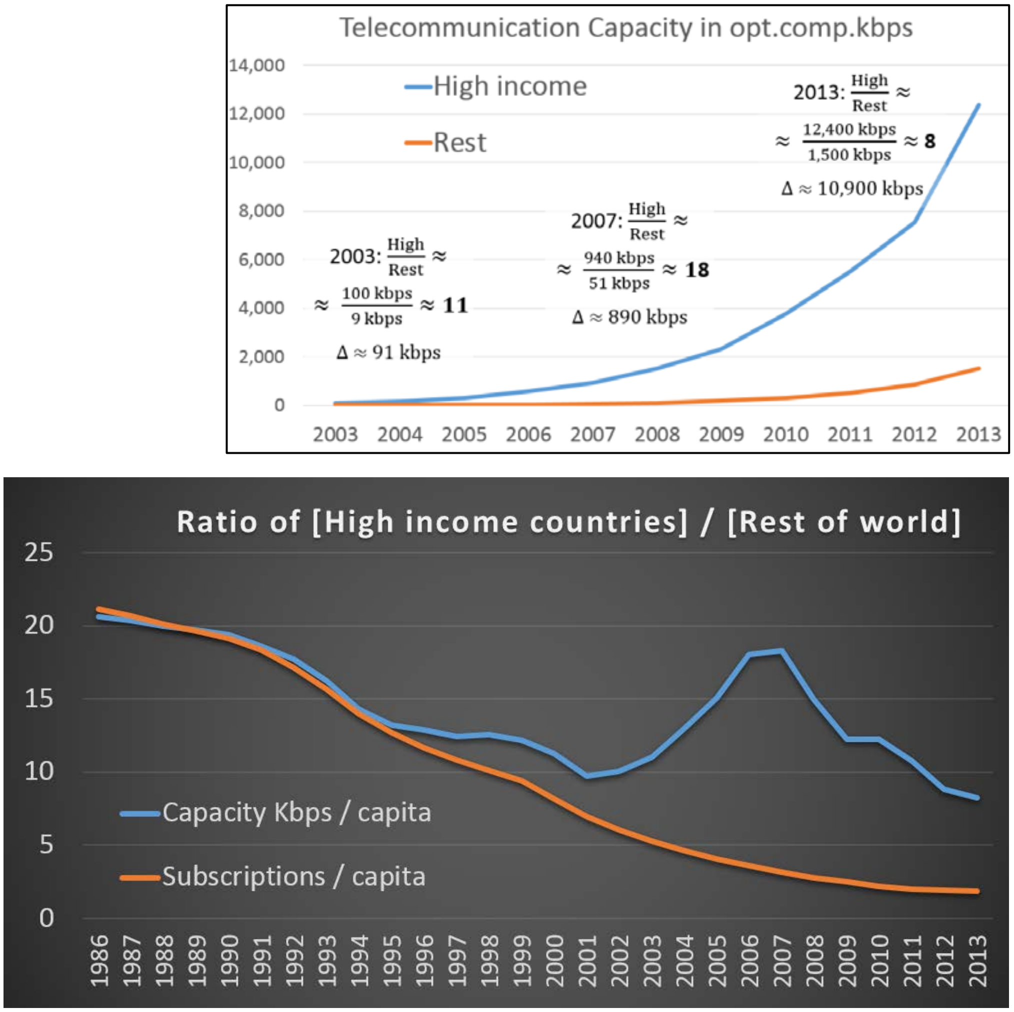
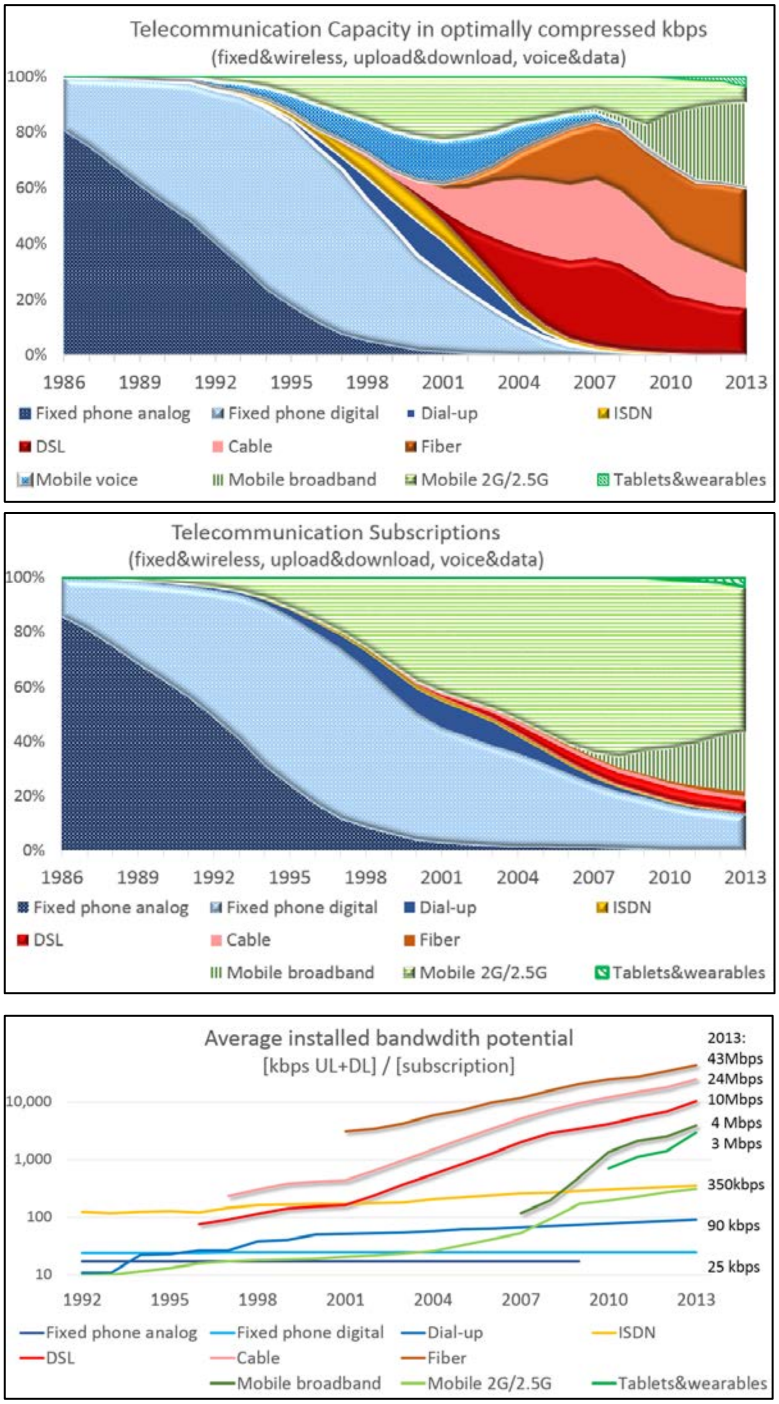


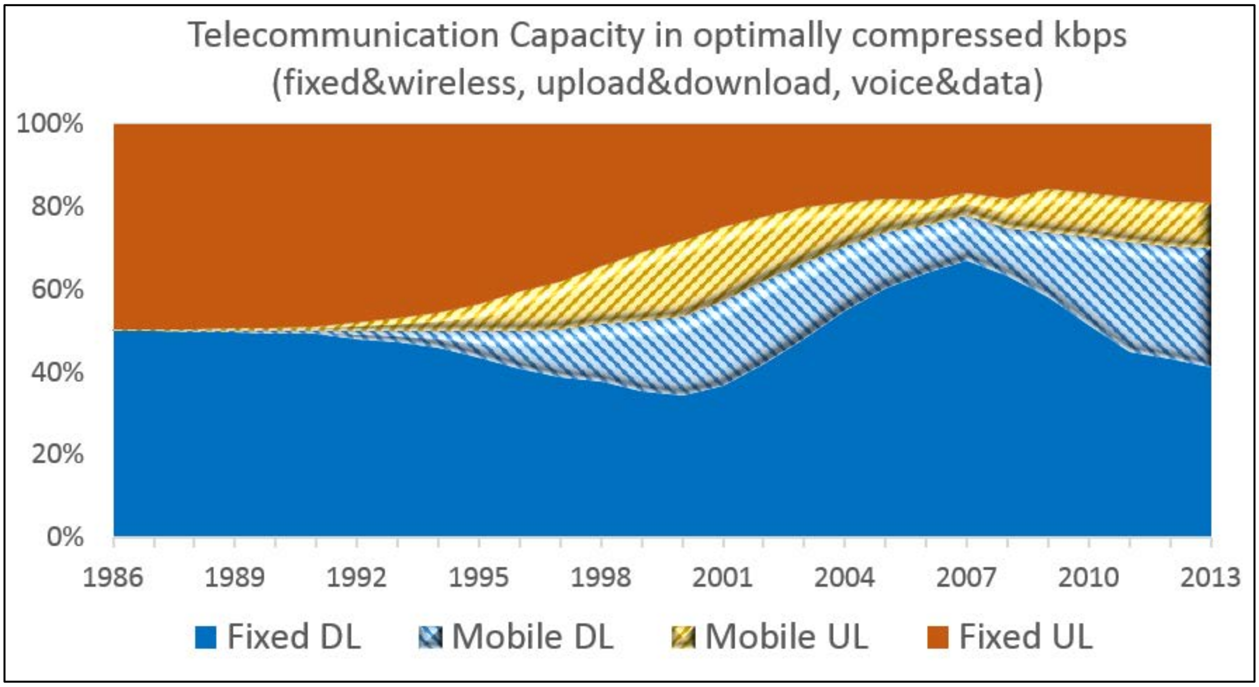
Figure 5 visualizes the source of the discrepancies between the tradition accounting of subscriptions, versus the accounting of Big Data potential through bandwidth potential measured in optimally compressed kbps. Both, the accounting of capacity (Figure 5a) and subscriptions (5b) evidence the elimination of the dominance of fixed line telephony, which was the dominating form of distance communication in the late 1980s. In terms of the number of technological devices, this dominance was clearly replaced by mobile telephony, especially narrowband 2G and 2.5G phones and in more recent years, broadband smart phones. However, in terms of telecommunication capacity, it shows that fixed-line broadband plays the dominant role. Representing less than 9 % of the world's subscriptions, it contributes 60 % to the global bandwidth potential. It can be expected that the share of effectively transmitted bits through this installed bandwidth potential leads to an even larger share of fixed-line broadband (for more in these methodological differences, see (Hilbert and López, 2012a). The increase of both mobile broadband in recent years is noticeable, as is the most recent contributions of tablets and wearables. Especially the latter have the potential to once again shift the picture of the global telecommunication landscape in the short-term future. Figure 3c shows the corresponding averages.

Figure 5: Technologies: (a) capacity kbps; (b) subscriptions; (c) kbps capacity per subscriptions



Finally, another aspect with important implications for the Big Data paradigm is the relation between uplink and downlink capacity. Uplink and downlinks show the potential of contribution and exploitation of the digital Big Data footprint. Figure 6 shows that the global telecommunication landscape has evolved from being a media of equal up- and downlink, toward to more download heavy medium. Up until 1997, global telecommunication bandwidth potential was equally split with 50 % uo- and 50 % down-link. The introduction of broadband and the gradual introduction of multimedia video and audio content changed this. In 2007, the installed uplink potential was a little as 22 %. The global diffusion of fiber optic cables seems to reverse this trend, reaching a share of 30 % uplink in 2013. It can be expected that the share of effectively transmitted bits through this installed bandwidth potential leads to an even larger share of fixed-line broadband (for more in these methodological differences, see (Hilbert and López, 2012a).

Figure 6: (a) Telecommunication capacity in optimally compressed kbps per uplink and downlink



hilbert@UCDavis.edu

References

Bauer, S., Clark, D., & Lehr, W. (2010). Understanding broadband speed measurements. Presented at the 38th Research Conference on Communication, Information, and Internet Policy (TPRC), Arlington, Virginia. http://people.csail.mit.edu/wlehr/Lehr-Papers_files/bauer_clark_lehr_2010.pdf

Bort, J. (2014, November 10). What is the "Race to Zero" and why does it make tech companies nervous?Slate.http://www.slate.com/blogs/business_insider/2014/11/10/cloud_storage_race_to_zero_means_companies_are_lowering_prices_and_raising.html

FCC (Federal Communications Commission). (2010). National Broadband Plan. http://www.fcc.gov/national-broadband-plan

Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big Data, Bigger Digi tal Shadows, and Biggest Growth in the Far East. IDC (International Data Corporation) sponsored by EMC. http://www.emc.com/leadership/digital-universe/iview/index.htm

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 Free access to the general public through: http://www.martinhilbert.net/worldinfocapacity-html/

Hilbert, M., & López, P. (2012a). 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.

Hilbert, M., & López, P. (2012b). 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.

Hilbert, M. (2012). How to Measure "How Much Information"? Theoretical, methodological, and statistical challenges for the social sciences. International Journal of Communication, 6 (Introduction to Special Section on "How to measure 'How-Much-Information'?"), 1042–1055.

Hilbert, M. (2014a). How much of the global information and communication explosion is driven by more, and how much by better technology? *Journal of the Association for Information Science and Technology*, 65(4), 856–861. doi:10.1002/asi.23031

Hilbert, M. (2014b). Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010. *Journal of the Association for Information Science and Technology*, 65(4), 821–835. doi:10.1002/asi.23020

Hilbert, M (2015a) Methodological and Statistical Background on a 2014 UPDATE to The World's Technological Capacity to Telecommunicate Information: http://www.martinhilbert.net/wp-content/uploads/2015/04/Methodological-Updates-2014.pdf

Hilbert, M (2015b) A review of large-scale 'how much information' inventories: variations, achievements and challenges. http://www.martinhilbert.net/a-review-of-large-scale-how-much-information-inventories/

ITU (International Telecommunication Union). (2012). Measuring the Information Society 2012. Geneva: International Telecommunication Union, ITU-D. http://www.itu.int/ITU-D/ict/publications/idi/

ITU (International Telecommunication Union). (2014). World Telecommunication/ICT Indicators Database. Geneva: International Telecommunication Union. http://www.itu.int/ITU-D/ict/statistics/

Kantei, (Prime Minister of Japan and his Cabinet). (2001). e-Japan Priority Policy Program. Government of Japan. http://www.kantei.go.jp/foreign/it/network/priority-all/index.html

López, P., & Hilbert, M. (2012). *Methodological and Statistical Background on The World's Technological Capacity to Store, Communicate, and Compute Information* (online document). http://www.martinhilbert.net/WorldInfoCapacity.html

Ookla. (2015). NetIndex source data. http://www.netindex.com/source-data/

Rhee, K. Y., & Kim, W.-B. (2004). The Adoption and Use of the Internet in South Korea. Journal of Computer-Mediated Communication, 9(4), 1208–1227.

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

Turner, V., Gantz, J., Reinsel, D., & Minton, S. (2014). The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things. IDC (International Data Corporation) sponsored by EMC. http://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm

hilbert@UCDavis.edu

Methodological and Statistical Background on a 2014 UPDATE to

Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65. https://doi.org/10.1126/science.1200970

Martin Hilbert
University of California, (www.martinhilbert.net)

This technical note describes changes and new data sources that have been introduced during the 2014 update of the 1986-2007 estimation of the world's technological capacity to telecommunicate[i]. If not otherwise indicated, the methodological choices and sources outlined in the original 300 page methodological supporting information still apply and should be cited i.e. López & Hilbert (2011; 2012)[ii]. The most important aspects are repeated in the following. The study includes 172 countries[iii], corresponding to 99 % of the world's Gross National Income (GNI, in current US$), and 96 % of the world's population.

1. Fixed-line phones

The assumptions are unchanged[ii]. Number of subscriptions is taken from ITU[iv] and performance metrics are replicated in the following table. Ley-μ is used in Australia, Japan and United States only.

Optimally compressed transmission rates for fixed line telephony.

		Analog	Digital
		"Optimal"	"Optimal"
Bit rate [kbps]	Ley-A	8.63	12.44
Bite rate [kbps]	Ley-μ Australia, Japan, U.S.	7.97	11.56

Source: López & Hilbert, 2011; 2012.

2. Mobile telephony

Until 2007 (and for 2G and 2.5G mobile), the assumptions follow López & Hilbert (2011; 2012)[ii]. We consider the differential start dates of data capacity in 2G phones (earliest 1992) and assume that all mobile phones have data capacity after 2002. In 2007, ITU[iv] stared to account for "active mobile-broadband subscriptions", which was completed and harmonized with previous estimates. Broadband bandwidth performance is constructed by comparing and complementing input from Ookla's Netindex[v] and Akamai[vi]. As usual, values for missing countries are estimated on basis of regional averages.

3. Broadband Internet

Most of the basic statistics for subscriptions are taken from ITU[iv], with complements and corrections as outlined in López & Hilbert (2011; 2012)[ii]. Performance estimations for dial-up and ISDN follow previous assumptions, as does broadband until 2007. From 2007 onward we use average bandwidth of countries reported by Ookla's NetIndex[v], which is seen "as the best of the currently available data sources for assessing the speed of ISP's broadband access service"[vii]. NetIndex compiles the results of two bandwidth velocity meters (Speedtest.net and Pingtest.net) and in this way estimates the average upstream and downstream speed for countries worldwide (e.g. for 2010 an average of 179,822 tests per country per day for 160 countries). We fill-in missing values on vasis of regional averages and countries with comparable profiles, and correct cases where we suspect a bias in the measurement.

The assignment of this total average toward specific technologies follows some assumptions (which do affect any analysis related to the contribudion of different broadband technologies, but not the total bandwdith, which stays the same). We assume that users of FTTH/B, DSL and cable modem (CM) perform most speed tests. We estimate the offered fiber optic bandwidth on basis of a large variety of national fiber optic providers, creating national averages among the offered rates. Based on several other speed-tests, we assume that cable modem download speed is 3 times faster than DSL, and cable modem upload speed is 1.5 times faster than DSL.[viii] Combined with the subscriber numbers, this gives us an equation with one unknown that can uniquely be solved for each country (in the following DSL performance for download):

ExcptValue [performance] =
= [% Fiber subsc.]*[Fiber perf.] + [% DSL subsc.]* [**DSL perf**.] + [% Cable subsc.]*3*[**DSL perf**.] =
= [National average bandwdith]

4. Tablets and Wearables

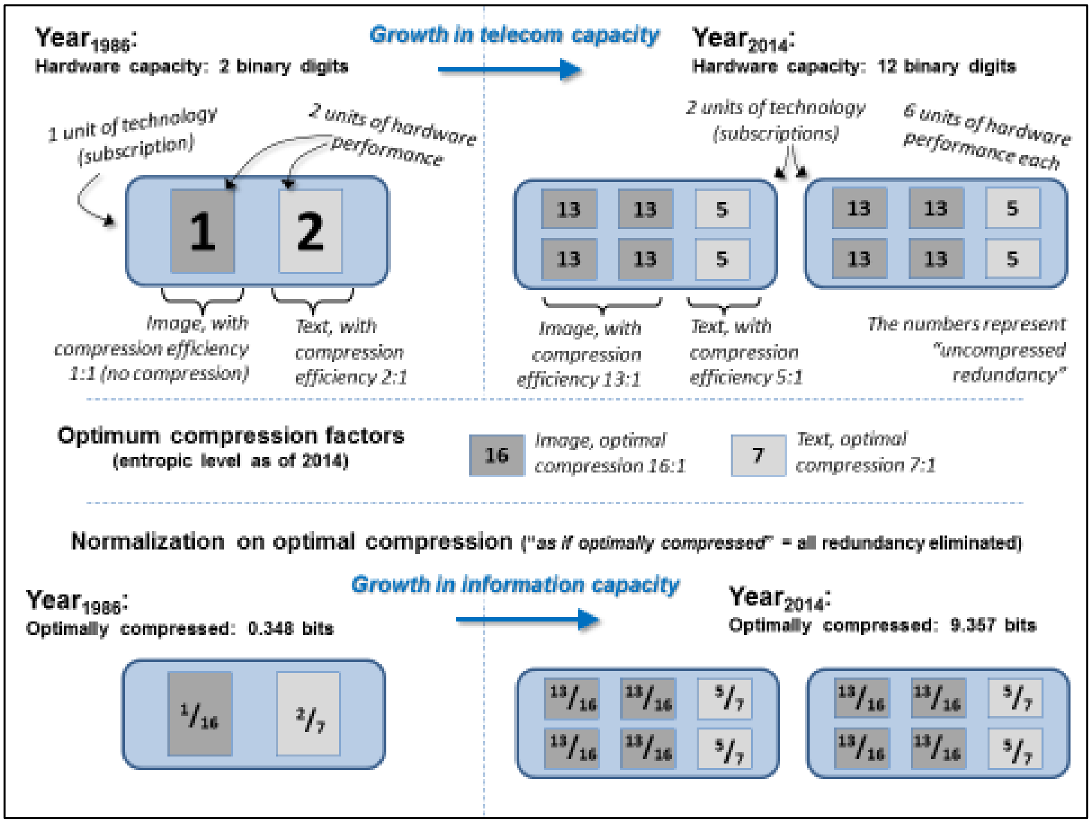
The number of tablets and wearables per world region is taken from the reports of Cisco[ix], Statista[x] and Ericsson[xi], while the number of per country is estimated on basis of the number of mobile phones per country (taking 2010 as the year of introduction of both tablets (i.e. the iPad) and wearables). After comparing the bandwidth for tablets and wearables presented in the above sources, it was decided to use the national 2G mobile phone bandwidth as an estimate for the bandwidth of wearables, and the mobile broadband bandwidth as an estimate for tablets.

5. Compression normalization

As explained in detail in Hilbert and López (2012)[xii] the estimation of time series makes it indispensable to work with some reasonable normalization on compression rates, as compression algorithms have enable to send more information through the same hardware infrastructure over recent decades[xiii]. We normalize on “optimally compressed bits” (“as if all content were compressed with the best compression algorithms possible in 2014”). It would also be possible to normalize on a different standard (e.g. the most used compression algorithms in 2014), but the optimal level of compression has a deeper information theoretic conceptualization as it approaches the entropy of the source)[xiv]. For the estimation of compression rates of different content, justifiable estimates are elaborated for 7-years intervals (1986, 1993, 2000, 2007, 2014). We create estimates for both compression factors and the respective content distribution for those years, and interpolate linearly between them.

The following Figure gives a schematic example of the procedure. Assume one subscription with a bandwidth of 2 bits/second in 1986. Half of this installed hardware capacity communicated uncompressed images and the other half text that is compressed with a factor of 2:1 (which means that uncompressed content is reduced to half of its size, for example through some kind of zip, rar, or related standard). In

2014, we count with two subscriptions (growth in subscriptions), each of which counts with a hardware bandwidth of 6 bits/sec (growth in hardware performance). Two thirds of this content is used to communicate images, and one third to communicate text. By now, images are compressed with a factor of 13, and text with a factor of 5. According to leading technological and theoretical considerations, the optimal achievable lossless compression for images is a factor of 16:1, and text 7:1. This is used to normalize the result on "optimally compressed bits" ("as if all content were always compressed with the best compression algorithms possible in 2014"). This quantifies the amount of information transmitted over the channel, not merely the hardware capacity.



6. Text:

In 2014, most text is still compressed with compression algorithms similar to the most common use of RAR in combination with ZIP. However, they evolved and RAR5 (introduced in 2013) achieves better compression rates that RAR3.7 used in 2007. Performance depends on file size and required speed, but we estimate an average improvement from 4.7 in 2007 to 4.9 in 2014.[xv] The optimal level of compression improved from a 6.6 to 7.2, which is achieved by algorithms like durilca' kingsize and cmix v6.[xvi]

7. Images:

JPEG has increased its standard as the dominant market leader (being also the standard for images on mobile phone services, a dominant source of images). We keep the previous assumption, but instead of a market share of GIF 36 % vs. JPEG 64 %, we assume a market share of 95 % JPEG. This leads to a compression factor of 37.7 for low quality for 2014 (from 27.6 in 2007; optimum stays 48), 22.8 for medium quality (from 17.6 in 2007; optimum stays 32), and 13.4 for high quality (from 11.3 in 2007;

optimum stays 16). We take a simple average for quality, and therefore get average compression factor of 18.8 for 2007; 24.6 for 2014; and a factor of 32:1 as optimal compression.

8. Sound:

For audio (including VoIP services) we recognize the increasing adoption of MPEG-4 AAC and AACv2, while the market share of MP3 is shrinking. We estimate an average compression rate of 20 for 2014, up from merely average 8.7 for average quality for audio voice in 2007 (6.8, 8.2 and 11.0 in low, medium, high quality audio/voice in 2007), and we adopt the optimum level of mobile communication of 32 for both mobile and fixed-line audio (as they have merged) (up from 24). For traditional telephony, during the period from 2007 to 2014 important improvements have been achieved for the quality of voice. However, this has not reduced size. We therefore use the same assumption for 2014 as we did in 2007.

9. Video:

H.265/MPEG-HEVC improved on H.264/MPEG-AVC during the period from 2007 to 2014. We follow a comparative assessment[xvii] and take the improvement of some 41 %, increasing compression from a factor of 60 to 85. As before, given the critical importance of video compression and the importance of centralized compression through streaming services like YouTube and Netflix, we assume that the optimal compression algorithm is also the most used one.

Basic compression factor per type of content

	1986	***1993***	***2000***	***2014***	***Optimal***
Text / compressed	2.2	2.9	4.6	4.9	**7.2**
Image	1.0	7.0	11.2	24.6	**32.0**
Sound	1.0	1.0	10.5	20.0	**32.0**
Video	1.0	20.0	27.0	85.0	**85.0**

10. Distribution of content:

The distribution of content is estimated on basis of a variety of sources, but mainly Sandvine (2014)[xviii].

Traffic distribution for services per region for 2014, based on Sandvine (2014)[xviii]

FIXED	Upstream		Downstream	MOBILE	Upstream		Downstream
North America							
BitTorrent	37%	Netflix	44%	Facebook	40%	YouTube	22%
HTTP	22%	YouTube	17%	SSL	19%	Facebook	17%
SSL	10%	HTTP	15%	HTTP	18%	HTTP	16%
Netflix	10%	iTunes	5%	YouTube	6%	MPEG	11%
YouTube	8%	SSL	4%	Instagram	5%	SSL	8%
Skype	3%	BitTorrent	4%	BitTorrent	3%	GoogleMarket	6%
Facebook	3%	MPEG	4%	MPEG	3%	PandoraRadio	6%
FaceTime	2%	Facebook	3%	PandoraRadio	2%	Netflix	6%
Dropbox	2%	AmazonVideo	2%	Gmail	2%	Instagram	4%
iTunes	2%	Hulu	2%	iCloud	2%	iTunes	4%
	100%		100%		100%		100%
Europe							
BitTorrent	45%	YouTube	26%	Facebook	27%	HTTP	25%
HTTP	14%	HTTP	24%	HTTP	20%	YouTube	23%
YouTube	10%	BitTorrent	15%	SSL	13%	Facebook	18%
SSL	8%	SSL	8%	YouTube	12%	SSL	8%
Skype	6%	Facebook	5%	BitTorrent	7%	MPEG	6%
Facebook	6%	RTMP	5%	Skype	7%	Netflix	5%
eDonkey	5%	MPEG	5%	iTunes	5%	iTunes	5%
Dropbox	3%	Netflix	4%	Instagram	3%	GoogleMarket	4%
MPEG	2%	FlashVideo	3%	MPEG	3%	BitTorrent	4%
iTunes	2%	iTunes	3%	Snapchat	3%	Instagram	3%
	100%		100%		100%		100%
Latin America							
BitTorrent	27%	YouTube	36%	Facebook	33%	Facebook	22%
YouTube	20%	HTTP	18%	SSL	17%	YouTube	18%
HTTP	17%	SSL	14%	BlackBerry	15%	HTTP	17%
Facebook	11%	BitTorrent	9%	HTTP	12%	BlackBerry	11%
SSL	11%	Facebook	7%	WhatsApp	8%	SSL	10%
Ares	5%	Netflix	6%	YouTube	4%	GoogleMarket	7%
MPEG	3%	MPEG	4%	Gmail	3%	Instagram	4%
Skype	2%	FlashVideo	3%	Twitter	3%	MPEG	4%
FlashVideo	2%	RTMP	2%	Ares	3%	WhatsApp	3%
Netflix	2%	GoogleMarket	2%	Skype	3%	Twitter	3%
	100%		100%		100%		100%
Asia Pacific							
BitTorrent	59%	YouTube	35%	HTTP	28%	HTTP	26%
QVoD	14%	BitTorrent	24%	Facebook	19%	YouTube	20%
YouTube	7%	HTTP	13%	SSL	19%	MPEG	15%
RTSP	5%	RTSP	8%	BitTorrent	9%	Facebook	10%
Thunder	4%	Facebook	5%	YouTube	6%	SSL	8%
HTTP	4%	MPEG	4%	Skype	5%	Dailymotion	5%
Skype	3%	QVoD	4%	MPEG	5%	GoogleMarket	4%
Facebook	2%	RTMP	2%	WhatsApp	4%	HTTPliveStream	4%
SSL	1%	FlashVideo	2%	Dropbox	3%	Instagram	4%
PPStream	1%	SSL	2%	Instagram	3%	iTunes	3%
	100%		100%		100%		100%
Africa							
BlackBerry	32%	HTTP	41%	BitTorrent	31%	HTTP	31%
HTTP	23%	BlackBerry	19%	HTTP	24%	YouTube	19%
WAPv2	13%	WAPv2	8%	YouTube	11%	BitTorrent	15%
SSL	9%	OperaMini	6%	SSL	11%	Facebook	9%
WhatsApp	7%	WhatsApp	5%	Facebook	9%	SSL	9%
Facebook	5%	SSL	5%	Skype	7%	MPEG	5%
OperaMini	4%	GoogleMarket	5%	MPEG	2%	FlashVideo	4%
BitTorrent	3%	YouTube	4%	FlashVideo	2%	Skype	3%
Skype	3%	Facebook	4%	iTunes	1%	iTunes	3%
Yahoo!Mail	3%	BitTorrent	2%	Dropbox	1%	GoogleMarket	2%
	100%		100%		100%		100%

Assumptions of 2014 content distribution for traffic flow services recorded by Sandvine (2014)[xviii]

	Text	Image	Sound	Video	Compressed
HTTPliveStreaming	0%	0%	2%	98%	0%
FlashVideo	0%	0%	2%	98%	0%
RTSP	0%	0%	2%	98%	0%
PPStream	0%	0%	2%	98%	0%
QVoD	0%	0%	2%	98%	0%
Netflix	0%	0%	5%	95%	0%
YouTube	0%	0%	5%	95%	0%
MPEG	0%	0%	5%	95%	0%
AmazonVideo	0%	0%	5%	95%	0%
Hulu	0%	0%	5%	95%	0%
BitTorrent	5%	3%	5%	85%	2%
Ares	5%	3%	5%	85%	2%
Dailymotion	5%	3%	5%	85%	2%
eDonkey	5%	3%	5%	85%	2%
Thunder	5%	3%	5%	85%	2%
Skype	0%	0%	30%	70%	0%
FaceTime	0%	0%	30%	70%	0%
Dropbox	15%	15%	15%	55%	0%
iTunes	1%	3%	50%	46%	0%
Snapchat	25%	30%	5%	40%	0%
WhatsApp	25%	30%	5%	40%	0%
RTMP	25%	30%	5%	40%	0%
HTTP	30%	31%	2%	35%	2%
BlackBerry	30%	31%	2%	35%	2%
Gmail	30%	31%	2%	35%	2%
Yahoo!Mail	30%	31%	2%	35%	2%
iCloud	30%	31%	2%	35%	2%
GoogleMarket	55%	10%	15%	20%	0%
Twitter	40%	35%	5%	20%	0%
Facebook	30%	40%	10%	20%	0%
WAPv2	60%	20%	10%	10%	0%
OperaMini	60%	20%	10%	10%	0%
SSL	70%	15%	5%	0%	10%
Instagram	5%	95%	0%	0%	0%
PandoraRadio	0%	0%	100%	0%	0%

The global content distribution that results from this methodology for 2014 is roughly reconfirmed by the global content estimates of Cisco Systems[ix], with the advantage that we now have a breakdown at the regional level (see Table).

Distribution of content type in % per region and Uplink / Downlink

FIXED						**MOBILE**					
North America DL	**1986**	**1993**	**2000**	**2007**	**2014**	***North America DL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	96	92	49	35	9	***Text/compressed***	100	100	100	40	20
Image	4	3	25	20	7	***Image***	0	0	0	26	18
Sound	0	5	19	6	7	***Sound***	0	0	0	6	14
Video	0	0	7	39	77	***Video***	0	0	0	27	48
North America UL	**1986**	**1993**	**2000**	**2007**	**2014**	***North America UL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	96	92	46	25	19	***Text/compressed***	100	100	100	40	32
Image	4	3	21	7	11	***Image***	0	0	0	26	31
Sound	0	5	27	12	7	***Sound***	0	0	0	6	8
Video	0	0	6	55	63	***Video***	0	0	0	27	26
Europe DL	**1986**	**1993**	**2000**	**2007**	**2014**	***Europe DL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	96	92	49	35	18	***Text/compressed***	100	100	100	40	22
Image	4	3	25	20	13	***Image***	0	0	0	26	19
Sound	0	5	19	6	6	***Sound***	0	0	0	6	8
Video	0	0	7	39	63	***Video***	0	0	0	27	51
Europe UL	**1986**	**1993**	**2000**	**2007**	**2014**	***Europe UL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	96	92	46	25	16	***Text/compressed***	100	100	100	40	26
Image	4	3	21	7	10	***Image***	0	0	0	26	23
Sound	0	5	27	12	7	***Sound***	0	0	0	6	9
Video	0	0	6	55	67	***Video***	0	0	0	27	41
Latin America DL	**1986**	**1993**	**2000**	**2007**	**2014**	***Latin America DL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	44	35	21	***Text/compressed***	100	100	100	40	30
Image	5	8	29	21	12	***Image***	0	0	0	26	26
Sound	0	6	20	5	5	***Sound***	0	0	0	6	6
Video	0	1	7	39	63	***Video***	0	0	0	27	39
Latin America UL	**1986**	**1993**	**2000**	**2007**	**2014**	***Latin America UL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	43	27	20	***Text/compressed***	100	100	100	40	36
Image	5	8	26	11	12	***Image***	0	0	0	26	28
Sound	0	6	24	10	6	***Sound***	0	0	0	6	6
Video	0	1	7	52	62	***Video***	0	0	0	27	29
AsiaPacific DL	**1986**	**1993**	**2000**	**2007**	**2014**	***AsiaPacific DL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	44	35	9	***Text/compressed***	100	100	100	40	21
Image	5	8	29	21	8	***Image***	0	0	0	26	18
Sound	0	6	20	5	4	***Sound***	0	0	0	6	6
Video	0	1	7	39	79	***Video***	0	0	0	27	55
AsiaPacific UL	**1986**	**1993**	**2000**	**2007**	**2014**	***AsiaPacific UL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	43	27	7	***Text/compressed***	100	100	100	40	32
Image	5	8	26	11	4	***Image***	0	0	0	26	24
Sound	0	6	24	10	5	***Sound***	0	0	0	6	7
Video	0	1	7	52	84	***Video***	0	0	0	27	38
Africa DL	**1986**	**1993**	**2000**	**2007**	**2014**	***Africa DL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	44	35	37	***Text/compressed***	100	100	100	40	22
Image	5	8	29	21	26	***Image***	0	0	0	26	15
Sound	0	6	20	5	5	***Sound***	0	0	0	6	7
Video	0	1	7	39	32	***Video***	0	0	0	27	56
Africa UL	**1986**	**1993**	**2000**	**2007**	**2014**	***Africa UL***	**1986**	**1993**	**2000**	**2007**	**2014**
Text/compressed	95	86	43	27	39	***Text/compressed***	100	100	100	40	21
Image	5	8	26	11	26	***Image***	0	0	0	26	14
Sound	0	6	24	10	5	***Sound***	0	0	0	6	7
Video	0	1	7	52	30	***Video***	0	0	0	27	58

hilbert@UCDavis.edu

---

[i] The original project covered the period 1986-2007 and resulted in a series of publications, among them:
Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://doi.org/10.1126/science.1200970
M. Hilbert, "What Is the Content of the World's Technologically Mediated Information and Communication Capacity: How Much Text, Image, Audio, and Video?," *The Information Society*, vol. 30, no. 2, pp. 127–143, 2014.
M. Hilbert, "How much information is there in the 'information society'?," *Significance*, vol. 9, no. 4, pp. 8–12, 2012.
M. Hilbert, "How much of the global information and communication explosion is driven by more, and how much by better technology?," *J Assn Inf Sci Tec*, vol. 65, no. 4, pp. 856–861, Apr. 2014.
M. Hilbert, "Technological information inequality as an incessantly moving target: The redistribution of information and communication capacities between 1986 and 2010," *J Assn Inf Sci Tec*, vol. 65, no. 4, pp. 821–835, Apr. 2014.

[ii] López, P., & Hilbert, M. (2011). Supporting Online Material: The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60 –65. http://doi.org/10.1126/science.1200970
López, P., & Hilbert, M. (2012). Methodological and Statistical Background on The World's Technological Capacity to Store, Communicate, and Compute Information (online document). Retrieved from http://www.martinhilbert.net/WorldInfoCapacity.html

[iii] Albania, Algeria, Andorra, Angola, Antigua and Barbuda, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahamas, Bahrain, Bangladesh, Barbados, Belarus, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Botswana, Brazil, Brunei Darussalam, Bulgaria, Burkina Faso, Burundi, Cameroon, Canada, Cape Verde, Central African Rep., Chad, Chile, China, Colombia, Comoros, Congo, Congo (Democratic Republic of the), Costa Rica, Cote d'Ivoire, Croatia, Cyprus, Czech Republic, Denmark, Djibouti, Dominica, Dominican Rep., Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Ethiopia, Fiji, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Greenland, Grenada, Guatemala, Guinea, Guyana, Honduras, Hong Kong, China, Hungary, Iceland, India, Indonesia, Iran (Islamic Rep. of), Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kiribati, Korea (Rep. of), Kuwait, Kyrgyzstan, Lao P.D.R., Latvia, Lebanon, Lesotho, Liberia, Liechtenstein, Lithuania, Luxembourg, Macao, China, Madagascar, Malawi, Malaysia, Mali, Malta, Mauritania, Mauritius, Mexico, Moldova, Mongolia, Morocco, Mozambique, Namibia, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Romania, Russia, Rwanda, Saint Kitts and Nevis, Saint Lucia, Samoa, Saudi Arabia, Senegal, Seychelles, Sierra Leone, Singapore, Slovak Republic, Slovenia, South Africa, Spain, Sri Lanka, St. Vincent and the Grenadines, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tajikistan, Tanzania, Thailand, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Turkmenistan, Uganda, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Uzbekistan, Vanuatu, Venezuela, Viet Nam, Yemen, Zambia, Zimbabwe,

[iv] ITU (International Telecommunication Union). (2014). World Telecommunication/ICT Indicators Database. Geneva: International Telecommunication Union. Retrieved from http://www.itu.int/ITU-D/ict/statistics/

[v] Ookla. (2015). NetIndex source data. Retrieved from http://www.netindex.com/source-data/

[vi] Akamai. (2015). The State of the Internet (Quarterly Report). Retrieved from http://www.akamai.com/stateoftheinternet/

[vii] Bauer, S., Clark, D., & Lehr, W. (2010). Understanding broadband speed measurements. Presented at the 38th Research Conference on Communication, Information, and Internet Policy (TPRC), Arlington, Virginia. http://people.csail.mit.edu/wlehr/Lehr-Papers_files/bauer_clark_lehr_2010.pdf

[viii] FTTH Council (2010). Rural Utilities Services Workshop - FTTH Deployment Status Update. October 2010, Retrieved November 2011, from http://www.ftthcouncil.org/en/content/ftth-council-presentations-to-rural-utilities-service-workshop-october-2010; http://s.ftthcouncil.org/files/rus_workshop_-_ftth_status.pdf
Organisation for Economic Co-Operation and Development (OECD). (2011). OECD communications outlook 2011 – Information and communications technologies. Retrieved November 2011, from: http://213.253.134.43/oecd/pdfs/browseit/9305011E.PDF
UltTex Speed Tester (2006). 3100 speed test results for 2005-2006 by type of technology. Retrieved November 2010, from: http://ult-tex.net/speed/results.pl?view_usa=All

hilbert@UCDavis.edu

[ix] Cisco Systems. (2015). Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2014–2019 White Paper. Retrieved from http://cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html

[x] Statista. (2014). Statistics and Market Data on Mobile Internet & Apps. Retrieved September 3, 2014, from www.statista.com/markets/424/topic/538/mobile-internet-apps/

[xi] Ericsson. (2015). Ericsson Mobility Report (OpenArticle). Retrieved from http://www.ericsson.com/mobility-report

[xii] 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.

[xiii] Hilbert, M. (2014). How much of the global information and communication explosion is driven by more, and how much by better technology? Journal of the Association for Information Science and Technology, 65(4), 856–861. doi:10.1002/asi.23031

[xiv] Shannon, C. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27, 379–423, 623–656. doi:10.1145/584091.584093

[xv] Softonic. (2013). WinRar 5.0 - is it really better than 7-zip? Retrieved April 12, 2015, from http://features.en.softonic.com/winrar-5-0-is-it-really-better-than-7-zip

[xvi] Mahoney, M. (2015). Large Text Compression Benchmark. Retrieved from http://mattmahoney.net/dc/text.html

[xvii] Grois, D., Marpe, D., Nguyen, T., & Hadar, O. (2014). Comparative assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders for low-delay video applications (Vol. 9217, p. 92170Q–92170Q–10). http://doi.org/10.1117/12.2073323

[xviii] Sandvine. (2014). 2014 Global broadband phenomena (Research Report). Sandvine Incorporated. Retrieved from https://www.sandvine.com/trends/global-internet-phenomena/