

ADVERTISEMENT

**SCIENTIFIC
AMERICAN™**

Sign In | Register 0

Search ScientificAmerican.com



Subscription Center

[Subscribe to All Access »](#)[Subscribe to Print »](#)[Give a Gift »](#)[View the Latest Issue »](#)[Subscribe](#)[News & Features](#)[Topics](#)[Blogs](#)[Videos & Podcasts](#)[Education](#)[Citizen Science](#)[SA Magazine](#)[SA Mind](#)[Books](#)[SA en español](#)[Technology » News](#)

4 ::

[Email ::](#)[Print](#)

More from Scientific American

Down in the Data Dumps: Researchers Inventory a World of Information

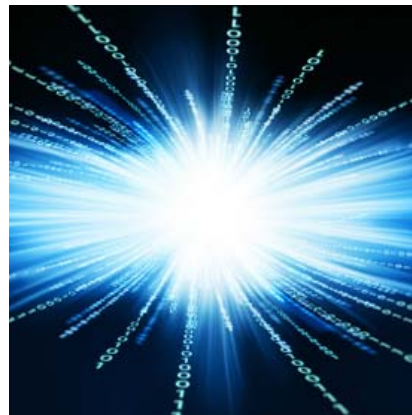
Nature still has the upper hand in terms of information storage and computational capacity, but this will not always be the case

February 10, 2011 | By [Larry Greenemeier](#)

Data are the common currency that unites all fields of science. As science progresses data proliferate, providing points of reference, revealing trends, and offering evidence to substantiate hypotheses. Decades into the digitization of science, however, data proliferate exponentially, at times threatening to drown knowledge and information in a sea of noise.

The journal *Science* [examines this trend in a special report](#) this week that, according to the editors, turns up two themes: "Most scientific disciplines are finding the data deluge to be extremely challenging, and tremendous opportunities can be realized if we can better organize and access the data." The report features among its articles an analysis of the challenges of understanding the reams of data being produced in particular by climate science, neurology and genomics.

One of the most interesting articles, however, attempts to quantify exactly how much data we're actually talking about and makes a key distinction between data and information. In "The World's Technological Capacity to Store, Communicate and Compute Information" [Martin Hilbert](#), a doctoral candidate the University of Southern California's Annenberg School for Communication and Journalism in Los Angeles, and Priscila López Pavez, a graduate student studying information and knowledge in society at the Open University of Catalonia in Santiago, Chile, report on their efforts to track 60 analog and digital technologies during the period from 1986 to 2007. The researchers found that the amount of data generated those two decades



COURTESY OF LOOPS7, VIA ISTOCKPHOTO.COM

MIND »



Classics »



DIGITAL »



Includes tablet edition!

ADVERTISEMENT



More to Explore

[Claude E. Shannon: Founder of Information Theory](#)[The Chemistry of Information Addiction](#)[Quantum Leap: Information Teleported between Ions at a Distance](#)

Latest News

Most Read

[Keep Impulses in Check By Looking at Nature](#)[Could an Animal's Name Save It from Extinction?](#)[Rebrand Stage Fright to Overcome It](#)[Sterile Neutrinos Still Theoretical](#)

exploded as digital technology moved into the mainstream. For example, the amount of data stored electronically in 2007 was equivalent to 61 CD-ROMs per person living on the planet at the time. For perspective, if those CDs were stacked, they would reach from Earth to the moon plus a quarter of that distance beyond.

Scientific American spoke with Hilbert, who also created and coordinated the [United Nations Regional Commission for Latin America and the Caribbean's Information Society Program](#) between 2000 and 2008, about the motivations behind his project with López Pavez, their research's potential impact on Hilbert's field of social science, and how human technological innovation stacks up against Mother Nature.

[An edited transcript of this interview follows.]

What prompted you to calculate the world's technological capacity to store, communicate and compute information?

In social science we've been talking about the digital revolution and the information society for quite some time now. We know that these technologies are the driver for productivity and the economy. We know that they're very important for political freedoms—just think about what's happening in Egypt right now. We know that they change the way a family is organized—consider how family members use cell phones to communicate while away from home. They change social conduct in every aspect. However, contrary to other sciences, social science does not yet walk the talk of the information age. Our paper is basically a contribution to bring social science toward the information age, which is important because information seems to be one of the unifying variables across all areas of science. One specific interest of ours was to see how fast information is growing and how fast we're digitizing this information.

There have been other studies that focused on measuring the hardware capacity of humankind. Now that's not information, that's just data. What we did here was normalize the compression rate, and that basically converts all data into information.

What does it mean to "normalize" compression?

The theory behind our research is actually pretty old and goes back to [U.S. mathematician, electronic engineer and cryptographer] [Claude Shannon's](#) information theory, which he introduced in 1948. So, basically what Shannon said is that we define information as the opposite of uncertainty. If you have uncertainty, you don't have information. And once you receive information, uncertainty is being resolved. He defines one bit as something that reduces uncertainty by half. We converted the data contained in storage and communication hardware capacity into informational bits. We measured information as if all redundancy were removed using the most efficient compression algorithms available in 2007.

In a practical sense, you can think of it like this: You have a Word document, and you save it on your hard disk. Let's say it's 100 kilobytes, and then you compress it with a zip file to only 50 kilobytes. What Shannon taught us is, if you compress it and compress it and compress it to the uttermost compression rate, we approach the entropy (or the actual amount of information) in this file. A compression algorithm takes out all of the redundant data in the file and leaves you with pure information.

In your research paper, you come across as pretty savvy with regard to the terminology and technology of information. Did you have to learn a great deal about the how data storage, compression, computation and other

Side Effects Vex Anti-Malaria Push in Sierra Leone

Follow Us: [f](#) [t](#) [g+](#) [v](#) [r](#)

See what we're tweeting about
Scientific American Contributors



tvjrennie RT @ClimateCentral: This is why renewables are at a turning point in the U.S. right now: <http://t.co/EUmMmGclJA>
19 minutes ago · reply · retweet · favorite



sciam SpaceX company hopes to launch the first ever reusable rocket this year [PODCAST]. <http://t.co/NxsEZM17OR>
#science <http://t.co/yZEWpObv9H>
1 hour ago · reply · retweet · favorite



sciam The wonderfully messy minds of creative people. <http://t.co/jOn17ktepO>
#science <http://t.co/1QHcMQy2FT>
2 hours ago · reply · retweet · favorite

[More >>](#)

Solve Innovation Challenges

DARPA Forecasting Chikungunya Challenge



Deadline: Feb 01 2015
Reward: See Details

DARPA(Defense Advanced Research Projects Agency) seeks methods to accurately forecast the spread of

Target Identification Mechanism for Bispecific Antibodies



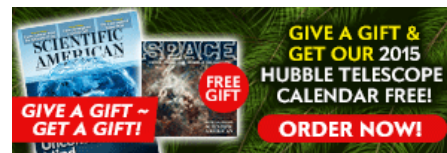
Deadline: Feb 06 2015
Reward: Varies

Bispecific antibodies (BsAbs) can offer advantages over conventional monoclonal antibodies (mAbs) because they have the potential to mod

[More Challenges >>](#)

Powered By: **INNOCENTIVE**

ADVERTISEMENT



Latest from SA Blog Network

Tadpole nests, past and present

Tetrapod Zoology | January 3, 2015

Physics Week in Review: January 3, 2015

Cocktail Party Physics | January 3, 2015

I'll Show You My Holiday Card If You Show Me

technologies work in order to pursue this project?

We had to learn a little bit. I'm an economist by training, and Priscila [López Pavez] is a telecommunications engineer, so I focused on the social statistics, the number of devices and the social interpretation of the information, whereas she focused more on the technology. Shannon was the one who taught us what information is and how to measure information. Our contribution was to take up this pretty old theory and convert it to a methodology that is useful for social science, and we applied this methodology for the first time to one concrete case—to measure how much information there is in the world, how much is stored, communicated and computed. This methodology could also be used for many other applications—for example, you could measure how much information there is in a company or in a tribe or in a society.

The methodology of your research aside, how does knowing the amount of the world's total technological capacity for storage, communication and computation help you in your role as a social scientist?

It'll help us a lot. If I say the total capacity for storage in 2007 was 295 exabytes, that's a huge number. [One exabyte equals one billion gigabytes, or one quintillion bytes.] Think about that amount of information this way: If you converted 295 exabytes of stored information into books, you could cover every square inch of the United States or China with 13 layers of those books. Yet this is still only about one third of the information that can be stored in all DNA molecules of a human.

We found that humankind received two zettabytes (one zettabyte equals one trillion gigabytes) of information through unidirectional broadcast in 2007, the equivalent of the amount of information each person would receive if they read 174 newspapers per day. Interestingly, however, the amount of information communicated in 2007 through bidirectional telecommunications, such as cell phones or e-mail was the equivalent of only six newspapers per day. This tells us that broadcasting is still massively superior to telecommunication in terms of the volume of information transmitted.

We also determined that computation is growing faster than either storage or communication capacity. That's interesting because when you hear people talk about the information society, people think about the Internet and mobile phones as a communication revolution. But actually it's kind of like the computation revolution because our computation capacity has grown twice as fast as our communication capacity. So it's really the computation that's a fascinating area that's often not appreciated enough because our focus goes to, you know, communication interfaces such as Facebook and Twitter.

You've compared the storage, communication and computation capabilities of technology with those found in nature—for example DNA storage. What did you learn from these comparisons?

Taking a look at our numbers, that's a vast amount of information, but if you compare them to nature they're still incredibly small. All of the DNA molecules in a single person can store 300 times more information than all of our combined technologies can store. The computational power of computers is extremely large, but the number of instructions they process per second is roughly the same as the number of neural nerve impulses that the human brain experiences in one second. Of course, I'm not trying to say the brain and a computer are the same thing, but this shows you how fine-tuned nature actually is. We say that we are pretty good with our technologies, and we're proud of them, but compared with what Mother Nature is doing, we are but

Yours.

MIND PsySociety | January 2, 2015

North American spadefoot toads and their incredible fast-metamorphosing, polymorphic tadpoles

Tetrapod Zoology | January 2, 2015

An American Pilot at War, 1915 (Part II)

STAFF Anecdotes from the Archive | January 2, 2015

News From Our Partners

REUTERS

Bid in New York to Extend Legal Rights to Chimps Fails, Again

nature

Side Effects Vex Anti-Malaria Push in Sierra Leone

techmedia

Jupiter Moon Europa's Giant Geysers Are Missing

ClimateWire

Major U.S. Cities Face More Blackouts under Climate Change

ADVERTISEMENT



ADVERTISEMENT

Airport Honolulu Hotel



Covered By
Expedia's Best Price
Guarantee
\$192.01

Science Jobs of the Week

Nanomedicine and Molecular Imaging
Northwestern University- Feinberg School of Medicine

Tutor in Physics
Massey University New Zealand

**Postdoctoral Research Associate in Theoretical
Condensed Matter Physics**
KTH Royal Institute of Technology

humble apprentices.

[More jobs from Naturejobs.com](#) »

Given the rate at which technology is advancing, what happens when it catches up with biology?

Now the difference between biological evolution and technological evolution is that, while biological evolution is incredibly powerful it is also incredibly slow. What will happen over the next century, we can be pretty sure, is that our technological capacity is coming on par with our biological capacity. You could estimate that by the end of the century all human brains combined could do as many nerve impulses as all of our computers combined could do instructions per second. And all our storage technology will store as much information as all human DNA can store. Some people refer to that as the singularity or whatever. I do not want to imply that a computer that does that many instructions is as intelligent or as smart as a brain, not at all. I don't think computers and people are the same thing, though they are surely complementary. Despite being humble apprentices now, we learn very fast. It is right around our generation and the generations to come that we'll get to the kind of complexity that nature is dealing with.

But you could also look at it the other way around. We spend \$3.5 trillion per year to improve the informational complexity of our technology, but what if we invested more in education? Less than \$50 is spent per child on their primary education in many parts of Africa. These numbers are a little out of sync, if you ask me. As a social scientist you ask: What would happen to social evolution if we would finally start to explore humankind's informational capacity?

The final year that your research covers is 2007. How might the numbers you calculated have changed in the past three years?

We covered more than 20 years of information and saw pretty stable growth rates over that time period. We saw that computation capabilities are doubling about every 18 months, so we could be pretty sure that since our study's inventory ended three years ago it has doubled two times already. Storage capacity doubles about every three years, so these 295 exabytes, multiplied by two, means right now we should have about 600 exabytes. We are pretty confident that you could extrapolate those figures easily today and also for the couple of years to come.

 Rights & Permissions

Share this Article:



Comments

Oldest - Newest

jtdwyer

February 10, 2011, 3:55 PM

An enormous amount of data storage capacity is allocated not to useful information but to replicated data, including copies of software program files and application installation and temporary data.

I assume that by far the majority of data storage capacity is configured on personal devices. Consider the allocation of storage on your own PC (you really have no idea how it's used, do you?). Excluding copied photos, videos, MP3s, backups, temporary files, etc. how much unique, original information is being stored? You know, data that's used in computational analyses to produce useful information... How much information do you compute, anyway?

It's absurd to consider the digital storage capacity installed on personal devices as 'data' in any

sense relating to 'information systems'.

[Report as Abuse](#) | [Link to This](#)

kristi276

February 13, 2011, 6:49 AM

In the Star Wars saga the question arises on the difference between knowledge and wisdom. The same could be said between the difference between information and knowledge, for information is the accumulation of the sum of knowledge spread over the course of the six million years of human experience quantified; whether, this knowledge is in a compressed or uncompressed state does not change the nature of the information (knowledge) accumulated. I believe that in order for information (knowledge) to be processed it first has to go through a series of quantitative, qualitative, and then to a quantum understanding of the knowledge (information) learned. In the Foundation, Harri Sheldon, the head of the Encyclopedist, devises a mathematical formula in order to save the accumulated knowledge of the Empire during the centuries of chaos.

We are in the infancy of the informational and technological age, and our collective knowledge has only begun to scratch the surface. The knowledge that will be accumulated in the next one hundred years will be far greater than the knowledge accumulated in the last one thousand years, and the rate of accumulation will increase exponentially over the course of the next thousand or five thousand years. How do you compress that?

[Report as Abuse](#) | [Link to This](#)

mo98

February 13, 2011, 9:48 AM

The permutative variations in data redundancy are alone enough to discredit any attempt at creating an ultimate compression strategy in information theory on any dynamic model of existence by simply adding or deleting copies of let's say, newspaper clippings in anyone's scrapbook. Absolute notions include the persistence of environment regardless of context. That said, the ultimate religion for artificial intelligence cannot be data and power supplies alone. Machines are dumb. Biology isn't. Both are adaptive and both can wear out and with that we redefine context in myriad ways. Food for thought.

[Report as Abuse](#) | [Link to This](#)

jstreet

February 13, 2011, 3:01 PM

It isn't clear to me that we humans will be capable of managing the huge amount of data into useful information before we become buried in it under a scientific Tower of Babel.

Two problems arise. One is the inability to find theories which unify the mountains of data into formulas linked together with logic and mathematics and the other is the inability to develop languages which make it possible for people in different fields to communicate with each other.

Scientists exist in small hermetically sealed enclaves who communicate among themselves with exclusive jargons which are incomprehensible to outsiders.

Why should we assume that the human species, which evolved to exist in small nomadic tribes of hunters and gatherers, can adapt to the glass walled modern "megalithic" cities most of us live in today?

We use languages with grammatical structures that are illogical accretions from their long pre-literate written history, and whose evolutionary laws and etiologies are unknown.

We can't even regularized English spelling, not to mention our irregular grammatical structures.

I'm not saying that we won't continue to make progress both socially and technologically, only that it is not a certainty. Maybe that is just another way of saying "God is dead" and there is no magic power above that will make things turn out well just because we humans say the right prayers and follow the rules of the day.

[Report as Abuse](#) | [Link to This](#)

You must [sign in](#) or [register](#) as a ScientificAmerican.com member to submit a comment.

TRY A RISK-FREE ISSUE



YES! Send me a free issue of Scientific American with no obligation to continue the subscription. If I like it, I will be billed for the one-year subscription.

[Subscribe Now](#)

Scientific American is a trademark of Scientific American, Inc., used with permission

© 2015 Scientific American, a Division of Nature America, Inc.

All Rights Reserved.

[Advertise](#)

[About Scientific American](#)

[Give a Gift](#)

[Special Ad Sections](#)

[Press Room](#)

[Renew Your Print Subscription](#)

[SA Custom Media and Partnerships](#)

[Site Map](#)

[Print Subscriber Customer Service](#)

[Science Jobs](#)

[Terms of Use](#)

[Buy Back Issues](#)

[Partner Network](#)

[Privacy Policy](#)

[FAQs](#)

[International Editions](#)

[Use of Cookies](#)

[Contact Us](#)

[Travel](#)