

BIG DATA: A SMALL GLIMPSE

Guy Philbin, Robert Morris University, philbin@rmu.edu

ABSTRACT

This paper posits that the growth of big data has paralleled Moore's Law, which stated that the number of transistors that could fit onto a silicon chip would double roughly once every two years. It presents a comprehensive definition of big data in a theoretical context grounded by Moore, Shannon, Weaver, Turing, and Ackoff; coins technological leaps and endpoints of an evolving roadmap that views big data as becoming; and suggests that future technology will move our understanding of big data from unintelligible symbols to valuable information that will eventually become knowledge.

Keywords: big data, Moore's Law, Information, Knowledge, Understanding, Wisdom

INTRODUCTION

This paper suggests that the growth of big data — like processing speed and memory capacity — has paralleled Moore's Law [19]. A key point regarding Moore's prediction is that each time an endpoint seemed inevitable, engineers achieved further innovations in miniaturization. The list of technological leaps in the history of silicon chips is long and includes innovations such as Dennard scaling, improved baking processes, high-k gate dielectrics, graphene carbon nanotubes, and other significant technological leaps.

If the recent history of science is any indication, more as yet unseen technological leaps in advanced materials, miniaturization, and storage innovation will continue to support the exponential growth of integrated circuits. For example, linear backbone polymers such as polyacetylene, polypyrrole, and polyaniline have resulted in a new class of organic materials offering the flexibility of plastics along with the electronic properties of traditional semiconductors. The full potential of organic molecules that can be chemically tailored to alter their electrical properties and other similar technological leaps have yet to be realized.

Data streams from humans and machines alike have turned into data rivers, which have turned into "exofloods" of information, that in turn flow into the oceans of information which have become a fixture of our lives and our technology. "This is big data," according to Gobble [8, p. 64], "and it is, many believe, the future." Nevertheless, as Moore [5] himself put it, the trend he predicted nearly 50 years ago is not likely to stop soon but, "It can't continue forever. The nature of exponentials is that you push them out and eventually disaster happens."

DEFINITION

Although clearly the data in our universe is expanding, as Mayer-Schönberger and Cukier [17, p. 6] have noted: "There is no rigorous definition of big data."

Early definitions of big data evolved from a realization that data and calculations were approaching or exceeding the physical limits of computer storage, memory, and processors. Unfortunately, definitions based on these kinds of parameters were unable to keep pace with the growth and scale of the phenomenon they were being used to describe.

Doug Laney [16], a vice president with Gartner Group, offered the "3V" definition of big data — volume, velocity, and variety — while he was with META Group in 2001. More recently, Intel's IT Center has expanded Laney's definition by adding value as a fourth useful characteristic. Here is Intel's [14] expanded definition:

- Volume. Huge data sets that are orders of magnitude larger than data managed in traditional storage and analytical solutions. Think petabytes instead of terabytes.

Issues in Information Systems
Volume 14, Issue 1, pp.162-168, 2013

- **Variety.** Heterogeneous, complex, and variable data, which are generated in formats as different as e-mail, social media, video, images, blogs, and sensor data—as well as “shadow data” such as access journals and Web search histories.
- **Velocity.** Data is generated as a constant stream with real-time queries for meaningful information to be served up on demand rather than batched.
- **Value.** Meaningful insights that deliver predictive analytics for future trends and patterns from deep, complex analysis based on machine learning, statistical modeling, and graph algorithms. These analytics go beyond the results of traditional business intelligence querying and reporting.

This “4V” definition is comprehensive and partly approaches the notion of big data as becoming. In particular, the notions of variety and velocity can be seen as a matrix of data streams in which both machines and humans generate information. Velocity is particularly helpful, as it can be defined as the speed of something in a given direction. If we understand big data in terms of fundamental quantities such as mass, length, or time raised to the appropriate power, we might envision big data as acceleration, for example, as $\text{length} \times \text{time}^{-2}$.

The notion of big data accelerating as the time rate of change of velocity in a certain direction ties in nicely with Moore’s Law and reducing a characteristic of this phenomenon to dimensions of physical quantities like change of velocity per unit time ($[\text{length}][\text{time}]^{-1}/[\text{time}]$) might make it possible to check the validity of various assumptions about big data, or even predict relationships that are not immediately apparent.

Big data that begins to describe the potential of data produced by intelligent systems and sensors and humans is a concept much like computer science itself, in that hypothesis and experiment are one. Extending this metaphor a bit, a program, loosely defined as a series of coded software instructions that controls the operation of a computer or other machine, works or does not work until conditions materialize for which it was not tested. The proof of the hypothesis — if it works — is contained within the code itself. Although others have defined big data with sheer size, processing capacity, digital exhaust, processing speed, clicks, unique links, or any other number of analytics imaginable, this paper would argue that big data is best conceptualized as potential.

IRREDUCIBLE UNIT

One of the most important elements of Moore’s Law [19, p. 33] is his notion of an irreducible unit: “By integrated electronics, I mean all the various technologies which are referred to as microelectronics today as well as any additional one that result in electronics functions supplied to the user as irreducible units.” Just as a grapheme is the smallest meaningful contrastive unit in a writing system, an endpoint will eventually come to the miniaturization of integrated electronic circuits. With apologies to Gordon Moore and linguistics, this significant technological endpoint might well be described as a moorepHEME. As with its linguistic counterpart, a moorepHEME would represent an irreducible unit of technology that cannot be further divided.

When this endpoint will be reached for integrated circuits is open for debate. Krauss and Starkman [15, p. 3], have suggested that the total amount of information that can be stored and processed has limits not dependent on its physical manifestations and that by implication, “Moore’s Law cannot continue unabated for more than 600 years for any technological civilization.” They arrived at this estimate in part by using the Bekenstein bound as an upper limit on information needed to fully describe a physical system to the quantum level.

In terms of computer science, the implication of this upper limit on information remains the subject of some debate. One argument holds that unbounded memory for a Turing machine is impossible because the finite size and energy of a given physical system imposes limits upon the rate at which information can be processed, although Turing [24, p. 437] himself saw no such limit on information: “There is no theoretical difficulty in the idea of a computer with an unlimited store.”

TURING TEST

Whatever the theoretical limits of machine memory and intelligence might be, Robert French in particular has asserted that big data and increased computational power and speed have made Turing and his test passé altogether and it is time to bid them adieu. As French [7, p. 76] framed this issue:

I am convinced no machine will pass a Turing Test, at least not in the foreseeable future, for the overriding reason I outlined earlier: There will remain recondite reaches of human cognition and physiognomy that will be able to serve as the basis for questions used to trip up any machine. So, set the Turing Test aside.

French concluded his argument to scrap Turing's Test by saying that we need to revise our current view of "understanding," regard computers as "high-level information sources," and replace our isomorphic human notions of cognition using "meaningful interactions" as the better measure of machine intelligence. While it is true that tasks requiring human intelligence were the central measures upon which the field of artificial intelligence was first grounded, this is not the place to argue whether or not the sapience of *Homo sapiens* — intelligence — needs be perfectly replicated by a machine. There is no small irony, however, in the fact that Turing [24, p. 433] created his test to avoid defining what is a machine and what it means to think.

As to intelligence, speculation that the internet is becoming an extension of human memory — much as encyclopedias served a similar role in the past — brings up the recurring possibility of a blurring between computer networks and human brains [12]. Lest this possibility seem unrealistic, Google's Chief Executive Larry Page [20] was famously quoted in a 2004 interview as saying, "On the more exciting front, you can imagine your brain being augmented by Google. For example you think about something and your cell phone could whisper the answer into your ear." Or perhaps even have a machine place an image before your eyes, or capture all you can see, which is what Google's 'smart glasses' promise to do [9], as yet another technology that will reap huge amounts of visual and personal data. To some, augmented reality glasses represents a golden harvest. To others, these glasses represent the end of privacy as we have known it.

In any event, the internet has already made inroads with respect to human memory. Researchers at Columbia, Harvard and the University of Wisconsin-Madison found what many of us already know all too well, namely that when people believe they can find information without committing it to memory they remember where to find it using what is called an external transactive memory resource [10]. In short, we have learned to "Google it." What this means in the long term for our mental actions — particularly how we store and acquire knowledge and understanding — remains to be seen [22].

BIG PICTURE

However cognition and understanding are measured, big data brings humans up against one of our most serious limitations when we compare ourselves with machines: information processing.

In *Strangers to Ourselves: Discovering the Adaptive Unconscious* Timothy Wilson [27, p. 24] suggested that the human mind can take in about eleven million pieces of information at a time, forty or so being conscious. This is an exciting but far cry from George Miller's *Magical Number Seven Plus or Minus Two* [18], a comparatively conservative estimation of the number of objects the average human being can hold in working memory. With no bias towards either approach to perception or cognitive processes — unconscious or conscious — there are clearly limits to human information processing which affect our ability to understand big data.

The titanic mass of big data already in the world today presents unique problems for human information processing. For example, without visual technology like geospatial information display, which allows us to "think in pictures," it would be difficult to comprehend twenty-four hours worth of airplane traffic over North America. Without visual representations of one sort or another — dependent upon data compression algorithms, intelligent agents, statistical models, expert systems, and so on — it would be impossible to make many real-time decisions based on big data.

The visual mapping and design of data allows us to see patterns and connections between and among numbers and bits of information that we would otherwise miss. Machines are needed to effectively deal with data streams

generated by events as diverse as thousands of airplanes in flight, hundreds of people responding to a natural disaster, or even a single human being streaming data during a bungee jump.

Yet visualization of big data must be done with some care as it compresses huge amounts of information into a very small space. Many graphic applications, by offering multiple ways to present data, make the data come alive. With each query, or iteration, or update, the picture changes and the underlying information grows. “And increasingly,” as Gobble [8, p. 64] has observed, “the biggest data is not created by humans.”

Of course, ophthalmoception is only one of the faculties that potentially allow humans to make big data decisions. Along with visual perception, humans can process data through hearing, taste, smell, and touch. Just as information technology input and output devices have evolved, visual encoding and decoding of structured and unstructured digital information is not the only means of interpreting, giving meaning to, and making decisions using big data.

Mayer-Schönberger and Cukier [17, p. 179] eloquently summarized these points when they stated that: “Big data operates at a scale that transcends our ordinary understanding.”

DATA

Today data often refers to facts and statistics collected together for reference or analysis. In the past, data was a term of philosophy that connoted what was known — or at least assumed to be fact — much as Diderot’s *Dictionnaire raisonne des sciences, des arts, et des metiers* was intended to gather, alphabetize, and present the whole field of enlightened human knowledge in an all-encompassing encyclopaedia [3]. Then as now data was the basis of reasoning and calculation. Such endeavors to comprehend all knowledge were long ago determined impossible and have been made even more so by the information age’s explosion of data, information, and knowledge. Doctorow [4, p. 17] has framed today’s situation in superb form regarding big data:

The mad, inconceivable growth of computer performance and data storage is changing science, knowledge, surveillance, freedom, literacy, the arts — everything that can be represented as data, or built on those representations.

Using Google’s Ngram Viewer, which currently allows a search of words and concepts like big data in about 500 billion words in 5 million books, below is a somewhat surprising result which indicates that as early as 1929 someone somewhere was already writing about big data. The rest of the story, as might be expected, shows the acceleration of interest in big data that began about 1955 and has continued to this day.

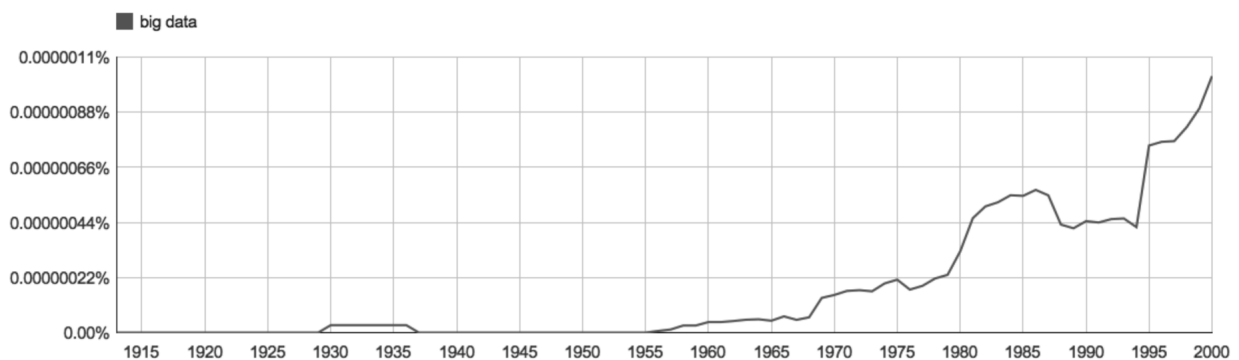


Figure 1: Big Data Ngram

Today data is used as a singular term when it means information and as a plural term in technical contexts where it means a collection of bits of information. From this frame of reference, big data may be viewed as something given, a premise from which rich inferences may be drawn. One of these inferences is that there is some indication that big data is by its very nature growing and becoming dynamic, rather than something static.

The dynamic elements of big data contain some essential commonalities. Yet these elements exhibit characteristics that are very different from one another. Big data contains both structured and unstructured data, which sometimes can be hard to define. In logic and mathematics, elements are single members of a set. In contrast, big data can be viewed as a mass of related information that satisfies clearly defined requirements, namely the quantities, characters, symbols, or patterns on which operations are performed by a computer. Rather than being individually specified, the elements of big data satisfy specified conditions.

In the context of information theory, big data seems to have no limitations in terms of a sequence, function, or sums of a series that produce a value. Rather, its boundaries are most often expressed in physical size, intensity of electrical signals, or recordings currently defined by magnetic, optical, mechanical, and other media (another Latin noun that no longer knows if it is singular or plural). Like the endpoint of Moore's Law, it is likely there is an upper limit to big data. However, as stated earlier, it is difficult to envision precisely when that will occur. This is because big data is literally a moving target and because it is likely there will be another information technology that will replace what it will become. That is, when theoretical technology becomes real. These and other technological leaps where experimental advances based on theories become reality might be described as theoreals.

UNCERTAINTY

No matter how it is defined, the big problem with big data is turning it into useful and actionable information. Ours is a world where there is far more information than knowledge. While this may seem trivial, the fact that our human capacity for understanding and processing information has limits is highly relevant to our understanding of big data.

A useful example is appropriate at this point. A large physical system, such as the weather, is an excellent metaphor for big data. Imagine the predictive accuracy and specificity that would be possible if every cell phone currently in use had an application that automatically reported barometric pressure, temperature, and location. If the data from such a massive information system were gathered and analyzed in real time, prediction of future weather events would move from hazarding a guess to making an astute inference based on actual meteorological data. Suddenly, everything from tracking microbursts to tornados becomes local and evidence-based. Real-time weather fed back to individual phones would complete the information loop (and perhaps motivate people to download and install the application). This general model has been applied to wide variety of situations apart from predicting if it will rain or snow. As Mayer-Schonberger and Cukier [17, pp. 11-12] framed the issue:

At its core, big data is about predictions. Though it is described as part of the branch of computer science called artificial intelligence, and more specifically, an area called machine learning, this characterization is misleading. Big data is not about trying to 'teach' a computer to 'think' like humans. Instead it's about applying math to huge quantities of data in order to infer probabilities...

The predictive accuracy of intelligent systems, defined as computer systems that mimic some aspect of human intelligence, is often based on statistical techniques grounded in Bayes' theorem, which is usually shown as: $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_i \{P(B_i)P(A|B_i)\}}$. While this formulation appears formidable, it is not. Expressed another way, this idea uses a set of "training data" to approximate the desired outcome and machine "learning" that occurs by taking advantage of the mathematical relationship between probabilities such that probabilities can be updated in light of new information. As Harper et. al. [11, p. 20] frame this issue of big data:

Nowadays, new techniques are being used to solve realtime inference problems in which multiple streams of data are processed from diverse sources. Statistical analyses are then used to make inferences about the state of the world. For example, when new information is received, probabilities can be updated using Bayes' theorem. This allows machines to learn by reducing the uncertainty of particular variables based on new information being fed into it.

Bayes nets and various computational tools that deal with uncertainty, conditional probability, and "messy" data have remarkable power and will undoubtedly continue to play a major role in finding logical and useful patterns in big data. Along the way, big data gets bigger when a smart information system takes initial data and uses it to update and refine relationships between probabilities in search of predictive accuracy. As the

system grows more accurate in its predictions, it also grows the base of data upon which it makes its predictions. The process of linking variables with predictive accuracy creates more data — and provides a ready example of how big data is becoming.

STATIC VS. DYNAMIC

Big data defies definition as simply a mass of zeros and ones in a linear sequence. Big data is dynamic by nature, and will likely change our understanding of information exchange, asymmetry, and processing in the bargain. What makes big data different is that it sometimes exhibits characteristics such as logic, consistency, and order, which make it coherent. In mathematics, coherent units are made up of only those units the quotient or product of any two of which yields the unit of the resultant quantity. While humans may not be able to see, or diagram, or even calculate how extremely complex variables relate to each other, machines can often provide meaning and show how things are connected.

If we use the spintronic chip as a metaphor, the brilliant mathematical information model Shannon [21] outlined nearly a half century ago is in need of a face lift. As a 2013 SciTechDaily [1] report on this technology framed it, a microchip that allows information to travel in three dimensions has great potential: “Researchers believe that in the future a 3D microchip would enable additional storage capacity on chips by allowing information to be spread across several layers instead of being compacted into one layer, as is currently the case.” This nanochip, which can store digital information as would a staircase of hard disk drives, allows data to move from left to right, front to back, and up to down.

If data can be viewed as a derived physical quantity capable of moving in two or three directions, this technology will surely translate into more and different information than Shannon envisioned in his linear mathematical model. In like manner, Warren Weaver’s [26] prescient suggestion to use statistical methods to translate natural language texts was ahead of its time in 1949. Computational power was simply not up to the task. Whereas today, great strides in computational power and cryptanalytic techniques have made this use of big data a reality in communication theory.

Just as information has to become knowledge before it is useful, big data is currently like a language that contains information which is useless without translation, a Tower of Babel overflowing with a multitude of languages. Understanding big data, appreciating just how important it is, represents a synthesis of a new kind of human knowledge. Taking what is previously known and synthesizing it with what is not known has often been the basis of incremental leaps in human development.

GAME CHANGER

Big data does not appear to be yet another information technology buzz word. While some, including Hunter [13, p. 56], have concluded that big data “...embodies an ambition to extract value from data, particularly for sales, marketing, and customer relations,” others view it as a bellwether of significant technological change. To frame it another way [13, p. 12]: “Big data is a game changer — and it’s already here.” Rather than a catchall for a vague idea that will go out of fashion, big data is a concept that captures a major shift in our technology driven world.

Conceptualizing the sheer volume, complex variety, real-time velocity, and predictive value of big data presents major challenges to our assumptions about how computations are made, stored and transmitted. Yet if our elementary view of data changes, so also will information theory. In the past, we have used metaphors like traffic for automobiles, telephones, airplanes, and networks. Looking to the future, matrix may be a more apropos figure of speech in describing an environment in which something develops, a matrix, medium, or material in which information flows and grows.

If the past is any predictor of the future, further miniaturization and improved performance of microelectronic components is likely to continue to an endpoint. The time immediately before Moore’s Law ceases to hold sway could be described as an information cliff that will signal the arrival of a moorepheme, or dramatically significant technological endpoint.

BIG BROTHER

Meanwhile, the massive increases in digital storage capacity in which the U.S. government and others are investing have opened some amazing possibilities. Edward Snowden, the former National Security Agency contractor whose revelations have created an international debate pitting individual privacy against national security, has become a symbol of these big data possibilities and pitfalls.

Snowden's revelations have drawn attention to a critical principle of big data: the more metadata, the greater its information value. Call detail records of a phone, for instance, can be used to analyze traffic patterns, time, place, frequency, duration, calls made and received, calls to others, associated phones, and the identity of partial telephone numbers. Pagers, cell phones, and computer networks add various means of location to this already impressive array of detail. To use a lower technology example, if the channel access protocol (CAP) of a pager is known, it is possible to see and record whatever messages are being sent to the pager.

Like the CAP of the pager world, there is a unique identifier that allows people to communicate on most computer networks called the MAC (media access control). The MAC provides a unique network address that internet providers use with dynamic host configuration protocol (DHCP) lease data to connect users. Because various manufacturers and service providers deal with metadata in nonuniform ways (there are different MAC configurations), availability and capture can vary.

The point is that massive amounts of metadata make it possible to ask questions that a decade ago would have been impossible. For example, in a letter to Darrell Issa (R-Calif.), chair of the House Committee on Government Reform and Oversight, Rep. Steve Stockman (R-Texas) [23] made this request:

As you know, recent revelations show the National Security Agency has been keeping an 'ongoing, daily' log of every domestic phone call in the United States. I respectfully request your Committee subpoena the records of every phone call made from all public and private telephones of all IRS personnel to all public and private telephones of all White House personnel.

In the unlikely event Stockman's request will be acted upon, as Russell Ackoff (1989) pointed out more than 25 years ago, without considerable machine sophistication finding and condensing the relevant information would be problematic. Simply having the requested metadata would not provide the desired information.

DATA TO WISDOM

In his foundational knowledge management paper which put forward the Data-Information-Knowledge-Understanding-Wisdom paradigm Ackoff defined data as symbols. Even 25 year ago, he aptly noted that system managers were "drowning in a sea of symbols," overloaded with meaningless data spewed out by mature information systems. This information overload which we now call big data began with an overabundance of symbols, which he described as "products of observations."

Two observations in Ackoff's freewheeling analysis *From Data to Wisdom* are especially relevant to big data, particularly the need for "less irrelevant information" and better "condensation of relevant information." He framed these related issues in terms of managerial efficiency, saying, "When everything is provided to managers already suffering from information overload, the amount of information they use decreases [2, p. 3]."

Another of Ackoff's [2, p. 9] intriguing observations was his description of the relationship between data and information, which he characterized as "functional and not structural," in that information is arithmetically or statistically inferred from data: "Information systems incorporate mental processes of relatively low order and therefore can to a large extent be automated. These systems generate information out of data."

Systems today employ advanced, automated, and highly complex algorithms to generate functional information, one of the essential values of big data. Turning again to Ackoff, functional information decays rapidly, knowledge becomes obsolete, understanding is not indefinite — only wisdom has permanence. Wisdom, for Ackoff, was “evaluated understanding” that provided a vision of the future. Moore’s Law represents a rare understanding of the future that has stood the test of time and provided us with a vision of big data. As Google’s Chief Economist Hal Varian (2012) more recently observed: “Data is so widely available and so strategically important that the scarce thing is the knowledge to extract wisdom from it [25].”

CONCLUSIONS

It seems quite likely that by the time the limits of Moore’s Law are reached and the miniaturization of integrated circuit technology comes to an endpoint, another more powerful information technology will displace whatever is then the most current circuitry. When the real leaps end and the size and power of integrated electronics reach their eventual limit we will have reached a significant technological endpoint that will most likely offer us a glimpse of the next phase of big data.

It is also likely the most exciting aspects of the relationship between human technology and big data have yet to be discovered. To use an analogy from the history of science, X-Strahlen (X-rays) were so named because their nature was unknown when they were accidentally discovered by Röntgen in 1895. We now know that high energy short wavelength electromagnetic radiation is able to pass through many materials, flowing in and out of bodies opaque to light. Although we cannot see or feel them, they have always existed.

Summarizing briefly, like much of big data — which flows by undetected — what we don’t know today and has yet to be discovered tomorrow may prove most useful. We are indebted to Moore for providing us with a handy metric to conceptualize the growth of big data and to Ackoff for pointing out that a system is the sum of the interactions of its parts. Future technology will doubtless move our understanding of big data from a growing mass of unintelligible symbols to very valuable information. Knowledge will reveal itself if we can frame the flow, connect the systems, and integrate the information that is passing by.

REFERENCES

1. 3D microchip allows information to travel in three dimensions. (2013, January) SciTechDaily. Retrieved from <http://scitechdaily.com/3d-microchip-allows-information-to-travel-in-three-dimensions/>
2. Ackoff, R. (1989). From data to wisdom. *Journal of Applied Systems Analysis*. 16. 3-9
3. Dictionnaire raisonne des sciences, des arts, et des metiers. [Systematic Dictionary of the Sciences, Arts and Crafts]. (1751-1772). Diderot, D. & d’Alembert, J. (Eds.). Available at <http://encyclopedia.uchicago.edu/>
4. Doctorow, C. (2008). Welcome to the Petacentre. *Nature*. 455(4). 16-21.
5. Dubash, M. (2005). Moore’s Law is dead, says Gordon Moore: Key predictor of IT will end sometime, reckons its progenitor. *Techworld*. April 13. Retrieved from <http://news.techworld.com/operating-systems/3477/moores-law-is-dead-says-gordon-moore/>
6. Elgin, B. (2004, May2). Google's goal: ‘Understand everything’ [interview] *Bloomberg Businessweek Magazine*. Retrieved from <http://www.businessweek.com/stories/2004-05-02/online-extra-googles-goal-understand-everything>
7. French, R. (2012). Moving beyond the Turing Test. *Communications of the ACM*. 55(12). 74-77.
8. Gobble, M. (2013). Big data: The next big thing in innovation. *Research-Technology Management*, January-February, 64-65.
9. Google glass features unveiled in preview video. (2013, February 20). *BBC News Technology*. Retrieved from <http://www.bbc.co.uk/news/technology-21519859>
10. Greenemeier, L. (2011, July 14). Piece of mind: Is the internet replacing our ability to remember? *Scientific American*. Retrieved from <http://www.scientificamerican.com/article.cfm?id=internet-transactive-memory>.
11. Harper, R., Rodden, T., Rogers, Y. & Sellen, A. (Eds.) (2008). *Being human: Human-computer interaction in the year 2020*. Cambridge, England: Microsoft Research. Ltd.
12. Hernandez, B. (2012, December 23). Google Working on Augmenting Human Brain. *NBC Bay Area*. Retrieved from <http://www.nbcbayarea.com/blogs/press-here/Google-Working-on-Augmenting-184504031.html>
13. Hunter, P. (2013). Journey to the centre of big data. *Engineering & Technology*. April. 56-59.

14. Intel. (2012, August). Distributed data mining and big data. Intel IT Center. Retrieved from www.intel.com/content/.../distributed-data-mining-paper.pdf.
15. Krauss, L. & Starkman, G. (2004, May). Universal limits on computation. Cornell University Library. Astrophysics. Retrieved from <http://arxiv.org/abs/astro-ph/0404510v2>.
16. Laney, D. (2001, February 6). Application delivery strategies. 3D data management: Controlling data volume, velocity, and variety. Current business conditions are pushing traditional data management principles to their limits, giving rise to novel, more formalized approaches. META Group. File 949.
17. Mayer-Schönberger, V. & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. New York: Houghton Mifflin Harcourt Publishing Company.
18. Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*. 63. 81-97.
19. Moore, G. (1965). Cramming more components onto integrated circuits. *Electronics*. 38(8).
20. Page, L. (2004, February 27). The Reuters story 'Google co-founder bugged by IPO speculation' is no longer available on CNET News. Copy available at <http://www.investmentbanter.com/showthread.php?t=44224>
21. Shannon, C. (1948, July, October). A mathematical theory of communication. *The Bell System Technical Journal*. 27. pp. 379–423, 623–656. Retrieved from <http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>
22. Sparrow, B, Liu, J. & Wegner, D. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*. 333(6043). 776-778.
23. Stockman, S. (2013, June). Press release. Retrieved from <http://stockman.house.gov/media-center/press-releases/stockman-requests-subpoena-of-nsa-s-white-house-irs-phone-logs>.
24. Turing, A. (1950). Computing machinery and intelligence. *Mind*. 59(236). 433-460.
25. Varian, H. (2012). Quoted in "Strategies For Setting Up Your Organisation For Success With Big Data." Kevin Long. Teradata. Retrieved from http://www.bigdatainsightgroup.com/site/sites/default/files/file_attach/Kevin%20Long%20-%20Big%20Data%20Strategies.pdf
26. Weaver, W. (1955). Translation (1949). In *Machine Translation of Languages*. MIT Press.
27. Wilson, T. (2004). Strangers to ourselves: Discovering the adaptive unconscious. Cambridge, Massachusetts: Harvard University Press.