

The data deluge: Five years on

In 2010 *The Economist* published “The data deluge”, a cover story by business correspondent Kenneth Cukier. It introduced the term “big data” to a general audience.

Since then the role of data in business and society has only grown. Data are the oil of the information economy; a vital corporate asset. The slides that follow show how the data deluge story has progressed since 2010.

For more:

<http://www.economist.com/node/15579717>



Data inflation

2

| Unit | Size | What it means |
|----------------|----------------------------|---|
| Bit (b) | 1 or 0 | Short for "binary digit", after the binary code (1 or 0) computers use to store and process data |
| Byte (B) | 8 bits | Enough information to create an English letter or number in computer code. It is the basic unit of computing |
| Kilobyte (KB) | 1,000, or 2^{10} , bytes | From "thousand" in Greek. One page of typed text is 2KB |
| Megabyte (MB) | 1,000KB; 2^{20} bytes | From "large" in Greek. The complete works of Shakespeare total 5MB. A typical pop song is about 4MB |
| Gigabyte (GB) | 1,000MB; 2^{30} bytes | From "giant" in Greek. A two-hour film can be compressed into 1-2GB |
| Terabyte (TB) | 1,000GB; 2^{40} bytes | From "monster" in Greek. All the catalogued books in America's Library of Congress total 15TB |
| Petabyte (PB) | 1,000TB; 2^{50} bytes | All letters delivered by America's postal service this year will amount to around 5PB. Google processes around 1PB every hour |
| Exabyte (EB) | 1,000PB; 2^{60} bytes | Equivalent to 10 billion copies of <i>The Economist</i> |
| Zettabyte (ZB) | 1,000EB; 2^{70} bytes | The total amount of information in existence this year is forecast to be around 1.2ZB |
| Yottabyte (YB) | 1,000ZB; 2^{80} bytes | Currently too big to imagine |

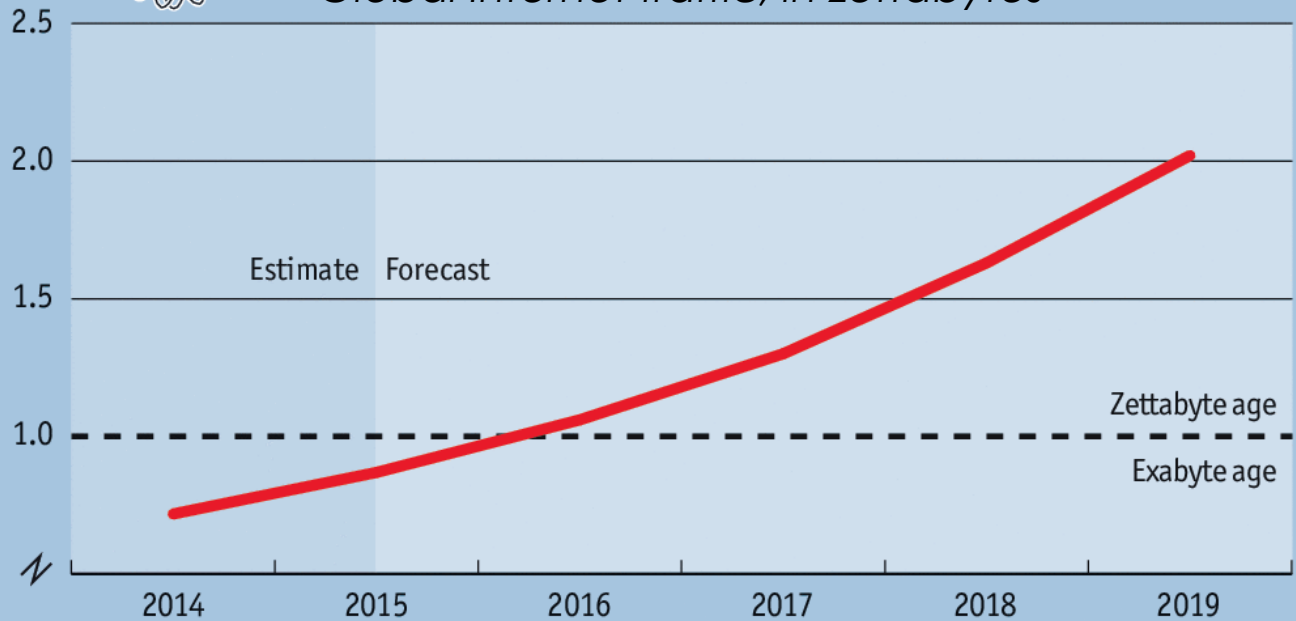
Source: *The Economist* The prefixes are set by an intergovernmental group, the International Bureau of Weights and Measures. Yotta and Zetta were added in 1991; terms for larger amounts have yet to be established.

The amount of data in the world is constantly expanding and doubles about every two years. Database pioneer Jim Gray liked to refer to terabytes as "terror bytes" – data so huge that they were scary.



Data with destiny

Global Internet traffic, in zettabytes



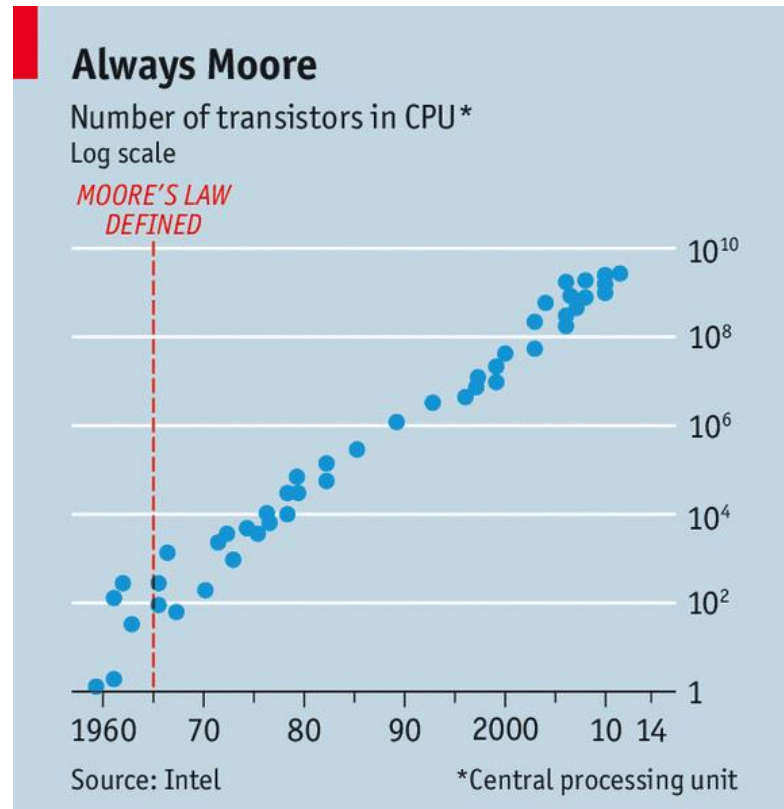
Source: Cisco

In 2015 the amount of data transmitted through the Internet reached about 870 exabytes of data. In 2016 we enter the “zettabyte age” as global Internet traffic rises above 1 zettabyte (about 1 trillion gigabytes).



The amount of data in storage, meanwhile, is predicted to reach 44 zettabytes by 2020. However, there are fears that the world's total data storage capacity will not be able to grow fast enough to meet demand.

Moore's Law, the idea that the power of a computer chip doubles every two years (as measured by the number of transistors), underpins the digital age. It celebrated its 50th birthday in 2015. But the law is starting to slow slightly – not because of technology but economics. The cost to produce ever smaller chips is increasing.



Economist.com

For more:

<http://www.economist.com/news/business/21648683-microchip-pioneers-prediction-has-bit-more-life-left-it-ever-more-moore>



Data are growing not just in volume but variety too. It is now estimated that 80% of data are unstructured, meaning they have no predefined data model. This presents new kinds of data management challenges: a plan by the city of Chicago to open its data to software developers requires cleaning up 10 billion lines of unstructured data.

The best disinfectant

Corruption and open data



Sources: Transparency International;
Open Knowledge Foundation

Economist.com

Many governments have made the data they compile accessible for public use – called the “open data movement”. There is a correlation between those countries that are most adept at making their data accessible and their lack of corruption.

For more:

<http://www.economist.com/news/international/21678833-open-data-revolution-has-not-lived-up-expectations-it-only-getting>

One of the areas most changed by data is human resources. Companies collect reams of data on their employees, from their CVs and performance reviews to their projects and locations.

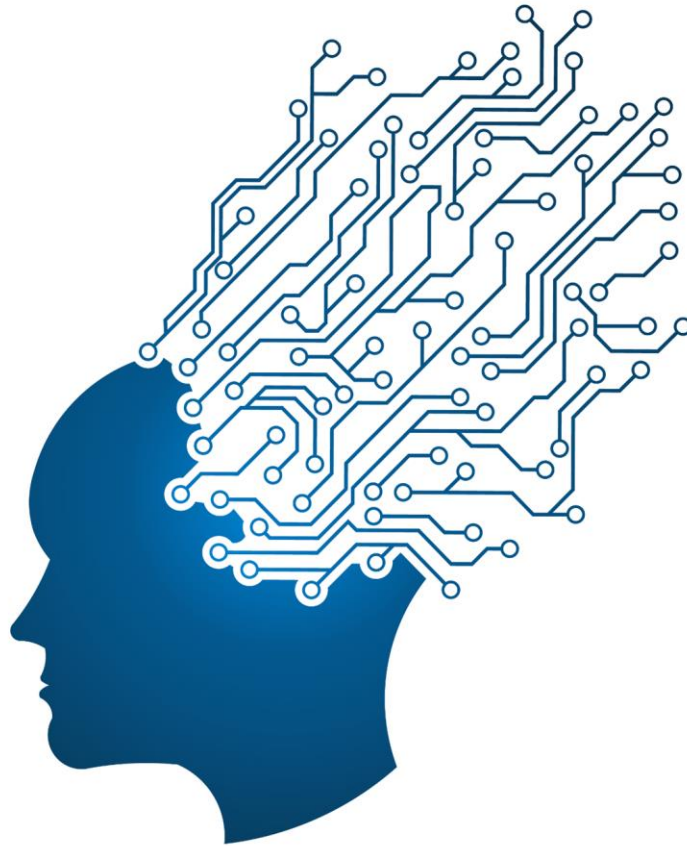
Firms use the data to improve recruitment and retention, which saves money. The data also help firms to place employees on teams to increase the odds that work is done on time and on budget.



For more:

<http://www.economist.com/news/business/21575820-how-software-helps-firms-hire-workers-more-efficiently-robot-recruiters>

Data power artificial intelligence (AI) algorithms and the technique of “machine learning”. The idea is not to explicitly programme a computer but to feed it enough data so it infers what to do, be it to drive a car or to detect cancer.



For more:

<http://www.economist.com/news/leaders/21650543-powerful-computers-will-reshape-humanitys-future-how-ensure-promise-outweighs>

Bring on the personal trainers

Probability that computerisation will lead to job losses within the next two decades, 2013
(1=certain)

| Job | Probability |
|-----------------------------|-------------|
| Recreational therapists | 0.003 |
| Dentists | 0.004 |
| Athletic trainers | 0.007 |
| Clergy | 0.008 |
| Chemical engineers | 0.02 |
| Editors | 0.06 |
| Firefighters | 0.17 |
| Actors | 0.37 |
| Health technologists | 0.40 |
| Economists | 0.43 |
| Commercial pilots | 0.55 |
| Machinists | 0.65 |
| Word processors and typists | 0.81 |
| Real estate sales agents | 0.86 |
| Technical writers | 0.89 |
| Retail salespersons | 0.92 |
| Accountants and auditors | 0.94 |
| Telemarketers | 0.99 |

Source: "The Future of Employment: How Susceptible are Jobs to Computerisation?" by C.Frey and M.Osborne (2013)

As more data are generated and processed by algorithms, many cognitive tasks now performed by human workers will be automated. According to a University of Oxford study, many of today's jobs are likely to be taken over by computers in the next two decades.

An understanding of what is possible with data and AI, and what remains the preserve of human intelligence, will be vital for both companies and workers in the coming decades.

For more:

<http://www.economist.com/news/briefing/21594264-previous-technological-innovation-has-always-delivered-more-long-run-employment-not-less>

In its leader article to accompany "The data deluge", *The Economist* wrote that "market incentives will ... come into play as organisations that manage data well are favoured over those that do not".

That process is still under way. But while the risks associated with the data deluge, including privacy, security and surveillance, are still with us, companies now recognise that mastering data is a vital component of success.

