Explainer: What is big data?

January 27 2014, by Asad Khan

So many people, so much data. Credit: Flickr/Fran Simó

Big Data, as the name implies, relates to very large sets of data collected through free or commercial services on the internet.

This massive amount of data arises from sensors, posts to social networking sites, digital images, videos posted online, transaction records of online purchases, and from mobile phone GPS signals to name a few.

A couple of aspects of big data worth noting are:

- it is impossible to remove/withdraw information from big data -

information once added will persist indefinitely in the cloud
- virtually any information that is stored electronically, including information within personal devices, offline data storage, even information thought to be deleted, has the potential to be included in big data.

A related development has been in sensory systems becoming online. Some as dedicated apparatus, others in secondary forms such as smart phones and tablet computers.

The unfolding landscape of numerous devices being connected to the internet – the Internet of Things (IoT) – will yield numerous personal and industrial applications such as internet-connected sensors for home automation, driver assistance, health monitoring, child and aged-care etc.

The transformation of big data into identifiable information has led to the development of open-access systems supporting forecast services. Vast information in the web when analysed as big data can be used to assess risk and increase competitiveness.

An area that has greatly benefited from big data analytics is demand-driven forecasting where decisions are formed from analysing huge volumes of data.

The potential for forecasting will continue to increase dramatically as location and other field data from sensors are included. As an example ambulance coordination systems that include weather and traffic forecasts will be more robust during critical periods.

But algorithms that will tap into the full potential of big data are not quite ready yet, particularly if sensory and device data are to be included within more conventional information systems such as online shopping and other web-based services.

Data mining and harnessing big data

Data mining is the process of analysing inter-data relationships – connecting the dots and finding hidden meanings and relationships that can provide startling new insights.
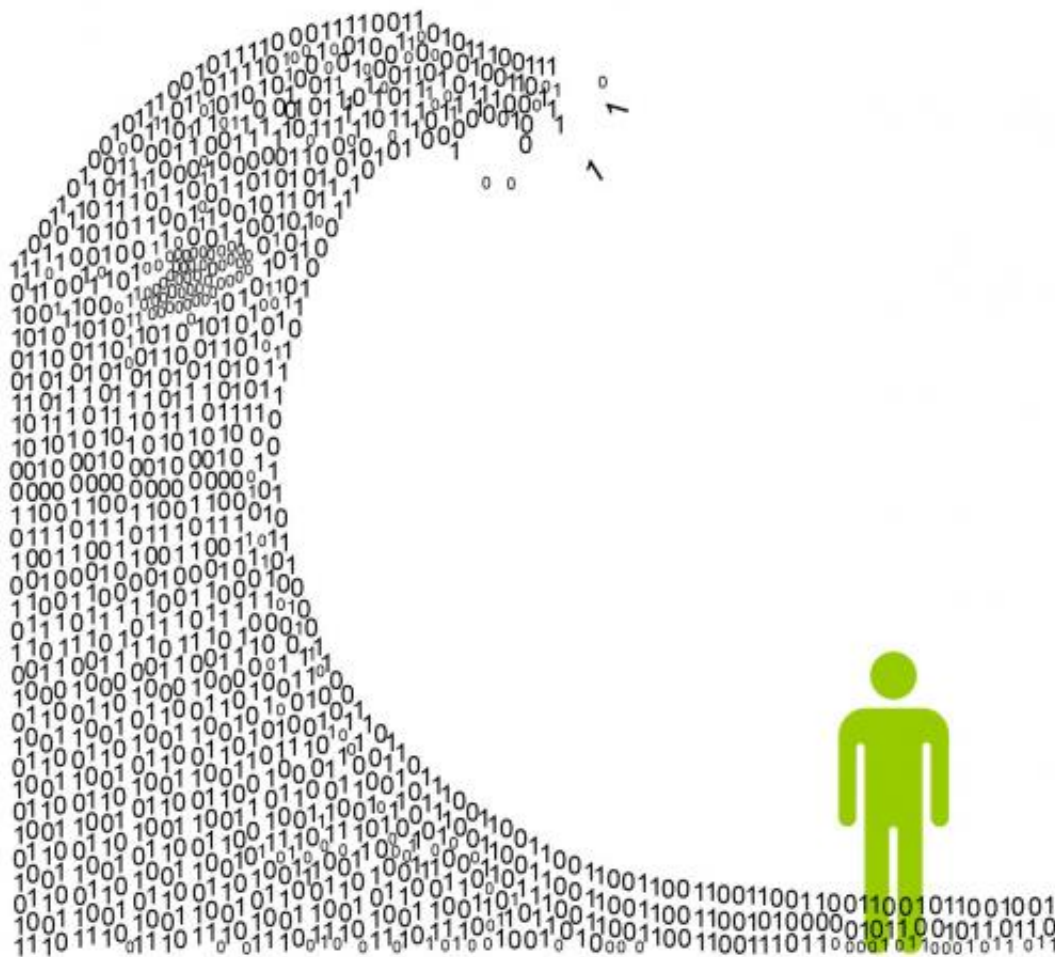
Credit: Intersection Consulting

This process of knowledge discovery provides information that can be used by industry to increase revenue and cut costs. This is where the

game changes from the cloud being merely a repository of vast information to a technology that yields considerable advantage to those who can properly utilise it.

Recent advances in parallel processing, distributed computing and high-performance computing (HPC) have enabled internet-scale data analytics that give strategic information to their operators.

What can be done with big data?

Consider the potential of being able to forecast the outcomes of certain types of world events or being able to answer specific questions relating to daily business matters.

Researchers in the UK studied 45 billion Google queries on a country-by-country basis and found that people in higher GDP countries show greater propensity in thinking of the future than people from lesser developed economies.

The competition to draw more accurate conclusions from universally-available big data on the internet is increasing.

Nature provides numerous examples of how it processes vast and disparate types of information sources:

- honeybees recognise fairly complex features in flowers: it has been shown they can even recognise human faces
- a fruit fly can conduct mind-boggling flight stunts with a miniature brain, small enough to fit on a pinhead, that uses minuscule amounts of energy.

We know the human brain has a far greater network density than any man-made network, yet it can quite efficiently integrate vast amounts of

[information](#) arising from inner processes and external sensations.

New studies from neuroscience are revealing details of the inner workings of the human brain. This has led to interesting [new algorithms](#), which tend to emulate brain functions for recognising sounds and patterns.

These computational models (known as bio-inspired or [biomimetic](#)) in principle should be able to interpret big data at internet-scale – which the brain does, with inner and sensory stimuli, at much higher scales.

But reproducing such processing within a conventional computer is extraordinarily time consuming. Emulating even a small part of the brain's activity for a very small period of time can take thousands of hours, if not days, on a desktop computer.

Presently a supercomputer is needed to simulate small [parts of the brain](#). Attaining the full potential of mining big data using brain-like processing, at this point in time, is not readily achievable.

Where to from here?

Australian universities have been at the forefront of analysing big data and presenting workable solutions, through projects such as [DART](#) and [ARCHER](#).

Their participation in international collaborative research in bio-inspired [wireless sensor networks](#) methods is opening new research paths that mimic the human brain in finding meaning through new forms of system designs, such as a [polymorphic computer](#) being research at Monash University.

Improvements to conventional computing methods are being researched

that will allow deeper interpretation of internet-scale data. The combination of new forms of computers that process data in [human brain](#) like manner, new algorithms derived from neuroscience, and advancements in cloud computing techniques provide a strong nexus for making strategic use of big data.

It may be argued that the capability to fully analyse internet-scale data will be key to nations in maintaining their prosperity and perhaps even security. The future may indeed rest with those with the best big data technologies.

This story is published courtesy of [The Conversation](#) *(under Creative Commons-Attribution/No derivatives).*

Source: The Conversation