6-2015

Big data: Big value and big concerns

Singapore Management University

BIG DATA: BIG VALUE AND BIG CONCERNS

Published:

29 Jun 2015

Digital information can serve lots of purposes, but timeliness, relevance and privacy issues abound

What is Big Data? IT research firm Gartner defines it as “high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making”.

“I see Big Data as collating data from disparate sources and putting them together to produce actionable insights,” muses **Anthony Bartolo** of internet service provider Tata Communications. “We’ve been making decisions for millennia with an almanac, and the almanac is the Big Data of yesterday. It fundamentally helped us decide when to plant crops, when to harvest etc.

“What’s happened between then and now is that the time frame (for decision-making) has been compressed. We now collect real time data, and before it becomes stale we can take action. “

Data for life

According to IBM, humankind generate 2.5 quintillion bytes of data – that is 2.5 million trillion – daily. Data that is relevant for immediate action, such as real-time adjustment of traffic light timing or bus scheduling, is processed to produce the necessary results before it does go stale. Data that is not needed immediately is stored, and given the falling cost of storage, it is often stored for eternity.

“We’re not storing for the sake of storing but we are storing for the future,” explains **Nils Michaelis** of Accenture Digital. Michaelis explains how the cost of one gigabyte (GB) of storage – over US$400,000 in 1980, now just 5 cents – has made it possible to keep data wholesale. In the past, he says, people stored only one percent of data when perhaps 35 percent of it is important. “Now it is economically viable to store 100% so we lose nothing,” he concludes.

Bartolo adds, “At one time, when the 1GB of storage space gets filled up, we’d have to throw out the old data and keep only the most recent data. Now there is no need to do that.”

That begs the question: Are we storing all the data just because it is affordable? Are we storing data that might serve no purpose at all?

“Data has a shelf life,” Bartolo says. “Some data are relevant in the short term, other data are relevant over the longer term. You cannot know for sure when a piece of data is going to be relevant and for what period of time until you understand what you are asking in the first place. Why throw anything out?

“Today, a piece of data might not be important because I might not be asking the right question, but I could ask a question years from now for which the data would be relevant. And it costs me next to nothing to keep it, so why not?”

He adds, “One of the fundamental underpinnings of analytics is sample size. The greater the sample size, the most accurate your reading of the trend will be. Compared to a reading of the data ten years ago, one done with a bigger sample size now would be more accurate.”

Singapore Management University Associate Professor **Archan Misra** of the School of Information Systems lists social media and big enterprises as the two main sources for Big Data. The former’s reputation for generating data is well known, but Misra describes the greater relevance of the latter.

“80 percent of our lives are lived in the real world: We queue for taxis, we queue for food, we scrap for seats on the train – these are all data we can harness to get a better understanding of individuals,” says Misra at the panel discussion, “*From Big Data to Smart Data: The future of wearable technology*” for Singapore-based station Channel NewsAsia’s Perspectives programme.

Such data, Misra says, can be crunched to produce better public policy. Favouring the term ‘Deep Data’ over ‘Big Data’, Misra says “it’s not just more microscopic details about your daily life in the physical world but the intrinsic value you get out of it. A lot of Deep Data today seems to be about extracting insights on the state of society but we haven’t made these insights valuable to individuals. That’s where I see possibilities, be they in insurance, retail, healthcare etc.”

Whose data?

Such insights, if they include a person’s health history, would also be valuable to insurance companies assessing the viability and health risk of someone signing up for a policy. In an era where data privacy is highly valued but less practised upon, consumers need to be aware of sharing their personal data.

“As consumers we make the decision on whether companies are successful or not in aggregating the data,” warns Bartolo, “and we make the decision whether to share the data. We as consumers have the right to exercise the right to deny companies permission to use our data. If we, as consumers, see the balance of power tilt in the wrong direction, we will force policy changes, or pressure governments to do so.”

What about data that one generates on social media? Would you be willing to pay a small fee to deny social media platforms access to your data, and in the process ensure privacy?

“The value of LinkedIn is the data,” says **Hari Krishnan** of professional network LinkedIn. “The only way the platform works is because everyone knows who everyone else is. You can see the people and the jobs you are interested in. If people don’t give us access to that data, it just doesn’t work that well.

“There is a cost-benefit analysis here but for LinkedIn, the less you give, the less you get out of it.”

The question, therefore, is: For what are you willing to give up your data?