Big Data—for better or worse

May 22 2013

Credit: AI-generated image ([disclaimer](#))

A full 90% of all the data in the world has been generated over the last two years. The internet companies are awash with data that can be grouped and utilised. Is this a good thing?

An increasing amount of data is becoming available on the internet. Each and every one of us is constantly producing and releasing data about ourselves. We do this either by moving around passively – our behaviour

being registered by cameras or card usage – or by logging onto our PCs and surfing the net.

The volumes of data make up what has been designated 'Big Data' – where data about individuals, groups and periods of time are combined into bigger groups or longer periods of time.

Research advantages

Petter Bae Brandtzæg of SINTEF ICT points to the huge research centres now developed at internet companies such as Facebook and [Google](link).

'The advantage they have is the enormous volume of data that other social researchers can only dream of,' he says.

However, it has also changed the way SINTEF researchers work. Even those not working in the major internet companies can still access Big Data.

Brandtzæg has investigated a tool called Wisdom developed by the American-based company MicroStrategy, and has started applying it in the delTA-project [deltaprosjektet.origo.no/](link) which addresses young people's social activity on the internet.

'This gives me access to data about over 20 million people – without making a single enquiry. I can analyse different preferences on Facebook and look at age and [gender differences](link) between various groups and nations across the world. So far I have compared gender differences in social activity on Facebook between people in Norway, Spain, England, USA, Russia, Egypt, India and China.'

Data protection is a problem we often associate with Big Data, but

according to Brandtzæg, data from Wisdom is restricted to large groups and does not go down to 'individual level'. This makes it possible for him to compare large groups without any data protection problems.

Short, transitory information

Big Data makes it possible to achieve research results that cover a wide range of issues, and can tell us a great deal about developments in the world in many different areas. It is possible to carry out thorough analyses and comparisons between countries and different genders.

For example, researchers in Facebook's own research department have looked into how people across the world update their messages, and what kind of information they post about themselves and their lives.

'The surveys show that the messages people have been posting have been getting shorter each year,' says Brandtzæg. 'This reflects the increase in other types of fast social communication, such as Twitter, which has achieved huge popularity because it is about expressing oneself briefly and concisely in a maximum of 140 characters. Another trend in that direction is that young people are telling their stories using images rather than text. The current Instragram craze could be due to the fact that you don't have to write anything.

Comparing data

These volumes of data can therefore provide us with useful information. However, Big Data can become a problem when different sources of data are compared for commercial use in targeted advertising campaigns.

It is becoming increasingly common for data about our location to be linked to our purchasing preferences – about what we like and don't like.

Facebook has made big strides in this area.

Vulnerability and data protection are the dark sides of our new entry into huge data sets and registers.

'Who knows – in two years, perhaps the tax register will be linked to the health and insurance register?' says Petter Bae Brandtzæg. 'And tax data can go astray; it has happened before.'

What opinions are being communicated?

The overwhelming volume of data being produced raises the issue of the content of all this information. What is being communicated?

The Networked Systems and Services department at SINTEF, to which Petter Bae Brandtzæg belongs, has recently had a bid accepted for the EU REVEAL project. In this project, researchers will look at combinations of different data sources and learn about people's ability to express themselves, and about the quality and truthfulness of data registered on social media. What is the content of these media? Who are the senders? Who else has said the same thing?

'We will look at various sources in relation to each other, and for example find out how trustworthy Twitter messages are,' says Brandtzæg. He also points to the new trend in fragmenting information across many channels – such as Facebook, SMS, e-mail, blogs, Twitter and Instagram.

How trustworthy are the media?

The ability to disseminate information to large groups in real time has made Twitter and Facebook important communications tools when

major events take place.

When hunting for the Boston terrorists, the police, authorities and traditional media also used social media like Twitter, Instagram, Reddit and [Facebook](#) to actively collect and disseminate information about the incident. Several voluntary groups were also set up via social media, in order to try and help the police. However, social media as channels of communications proved to be not entirely beneficial, but also a source of confusion and misinformation.

Can Big Data be used as a resource for journalists, and how trustworthy is the information available on social media? This is one of the subjects that the SINTEF researchers will be looking into as part of the EU REVEAL project.

Provided by SINTEF