

THE HACK'S TOOL BOX



Hacks and hackers

A new era of 'data-driven journalism' is changing the nature of reporting

By Kenneth Cukier

The state of technology affects how people get their news. After Gutenberg invented printing in the 15th century there was an explosion of information. It took the creation of a new intermediary between the information and the public – which we now call a journalist – to make sense of it.

Today, a similar shift is taking place. Digital technologies are producing a deluge of data that previously didn't exist. The quantity of information doubles every three years, according to one measure. As a result, a new branch of reporting is emerging: data journalism.

Instead of news stories based on a string of anecdotes and discrete slivers of facts, the data journalist is interrogating massive databases to uncover new insights and scoops hidden in plain sight – that is, buried within the bits and bytes.

Sometimes the output is a written article. More frequently, however, the story is best communicated visually, as elaborate or interactive infographics. Data journalism won't replace the time-honored ways of media. But it is an important new tool with which the media can explain the world to its audience.

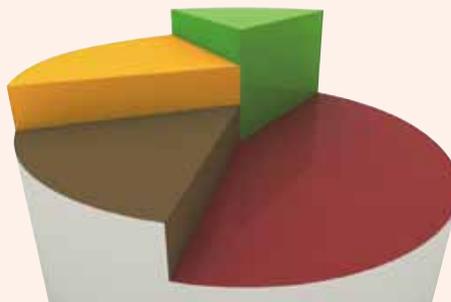
A few examples illustrate what's happening. In 2010 the *Guardian* combed through the Wikileaks release of confidential U.S. government cables. It compiled 91,000 military reports from the Afghan War to create an interactive map of the date, location and casualties from 16,000 "improvised explosive device" (IED) attacks. It was a way for people to judge the effectiveness of America's military presence in the country, but presented in a form that told the story more effectively – and in some ways more rigorously and accurately – than could be presented in words after inter-

viewing people. The journalists let the data tell the story, not the sources.

That case is special in part because the data source was classified – our generation's "Pentagon Papers." More commonly the data comes from national statistical offices or international organizations that would like nothing better than for the news media to take their underlying data and produce something new with it. You can appreciate their frustration.

After spending years deep in the data and giving birth to a 200-page report, the media and policy makers normally pick out one or two tidbits and a day later are on to something else. The vast majority of information gets lost in the noise pollution of daily journalism. As a result, the public, and the topic itself, gets short shrift. Although some official bodies dislike the transparency and try to keep their data to themselves, many more welcome the newfound interest in their work.

Data surrounds us everywhere, even when it's scarcely considered data at all. But with a little imagination the information can be presented in visual way to say something novel. For instance, when Facebook was suffering one of its many privacy conflagrations a year ago, the data desk at the *New York Times* had the clever idea of counting the frequency of changes and number of words in the company's privacy policy, plotted over time.



The result was brilliant. One clearly sees that as Facebook grew, it made more changes: a policy that started with a few hundred words soared into the tens of thousands. No one had ever considered a legal disclaimer as "data" – but the infographic told the story better than the article did. (Likewise, *The Economist's* Big Mac Index, in which the price of a burger in many countries is used to compare purchasing power parity, is an example of playfully using figures to tell a story.)

Even a single fact or two, spotted by crunching the numbers or thinking about information in a new way, can be the basis of a major article. Regardless of one's view of America's recent wars, it is hard not to be stunned upon learning that more American soldiers died from suicide than in combat over the past two years.

A problem with data-driven journalism is that sometimes the data is as unreliable an interlocutor as any source. The information might be biased, incomplete or interpreted by the journalist incorrectly. Take, for instance, the above point about soldiers and suicide. To some, it suggests the wars were a tragic misadventure. But another interpretation is that the wars are being won; that there are far fewer combat deaths nowadays.

All reporting is based on information. But parsing databases requires new skills. Journalism needs to change and adopt new tools. In the past, the press was quite literally the press. Yet today's hacks have more in common with hackers. It marks a fascinating evolution for media, and how news gets provided to the public. ❶

Kenneth Cukier was *The Economist's* Japan business correspondent from 2007 to 2012. In March he returns to London as the paper's first "Data Editor".