

## BIG DATA

# The revolution is digitized

Charles Seife digs into three studies of the wild new world of big data.

The term has been around for almost two decades, but the world only really started talking about 'big data' in the first few months of 2011. We know this because we can look it up on Google Trends (see 'Byte marks').

Google built its empire on gathering and analysing nearly unfathomable depths of data. Every query ever typed into the search engine is sitting in Google's multi-exabyte data stores. These stores also hold the full text of tens of millions of books, high-resolution images of streets around the world and myriad e-mails, videos, word-processing documents and spreadsheets. Anything that can be rendered in bits and bytes and is accessible to the company's servers will be pushed, filed, stamped, indexed, briefed, debriefed and numbered by semi-autonomous information-gathering agents. Enter 'big data' into the Google Trends website and, a fraction of a second later, a graph of frequency appears, its line rising sharply upwards in the first quarter of 2011. You are distilling that information from a colossal data set containing the entire world's search-engine queries for the past ten years.

Seamless upgrades in computer interfaces masked a liminal moment: in a few years, we have moved from data that can be created, gathered and understood by unaided humans — kilobytes, megabytes and gigabytes — into the hitherto unimaginable realm of petabytes and exabytes, gathered at terahertz speeds and processed almost as quickly. The transition has moved beyond scale to revolution.

In *Big Data, Little Data, No Data*, information-studies specialist Christine Borgman looks at big data through a fairly narrow lens: academic research. Each day, scientists grapple with ever more appalling volumes of data. The ATLAS detector on the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, has to sort through dozens of terabytes of data every second while it is running — and filter that down by five orders of magnitude before humans can deal with it. Next-generation

## Big Data, Little Data, No Data: Scholarship in the Networked World

CHRISTINE L. BORGMAN  
MIT Press: 2015.

## Data-ism: The Revolution Transforming Decision Making, Consumer Behavior and Almost Everything Else

STEVE LOHR  
Harper Business: 2015.

## Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World

BRUCE SCHNEIER  
W. W. Norton: 2015.

telescopes such as the Square Kilometre Array will be gathering exabytes of data each day — an amount that would have filled the total storage capacity of all the world's information-carrying devices (including books, photos and videos) up to the mid-1980s.

Borgman is something of a data anthropologist. She goes among researchers in physical sciences, social sciences and humanities alike to find out how they collect, handle and share the flood of information. Her treatise is interesting, but frustrating. She has difficulty turning her sizeable data set into a narrative both broad enough to cover the range of topics and deep enough to do justice to them. All too often, she seems to give a quick nod to essential elements. For example, she mentions open publications and data, but provides no hint of the battles around them in the research and publishing worlds. She offers key insights — that there are different dynamics to publishing research results and raw data, and that it is shortsighted to focus on releasing new data sets rather than on how to preserve and reuse the data. But the book might have said much more. There is nary a word about the huge controversy around incursions of commercial entities into the gathering, dissemination and control of scholarly data.

In *Data-ism*, Steve Lohr goes after the commercial implications of big data, but through an equally narrow lens. As a veteran technology and business reporter, he is attracted to the story of how data can help to root out

## BYTE MARKS

The surge of interest in big data since 2011 can be clearly traced in Google's archive of search terms.

SOURCE: GOOGLE TRENDS; GRAPHIC: CLAIRE WELSH/NATURE

2005

2007

2009

2011

2013

2015

inefficiencies that stop businesses reaching their potential. He gives the example of McKesson, a drug and medical-supply distributor that used its archives of product and shipping data to create a supply-chain model. That led to a billion-dollar decrease in inventories and a sizeable jump in efficiency, showing, as Lohr says, “data really being used to ... make better decisions, ones that trump best guesses and gut feel, experience and intuition”.

Alas, *Data-ism* is very much a conventional business book, full of anecdotes, mini-profiles and aphorisms that grow ever less compelling, however well they would go over at a TEDx talk. Lohr’s journalistic instincts often seem to betray him. He is unimpressed with the massive data-collecting and consumer-profiling of information giant Acxiom, yet bowled over by a seemingly conventional personality-horoscope program that snaffled up Twitter feeds, and, for 81% of subjects, “pretty much matched the results of their formal tests for personality type, basic values, and needs”.

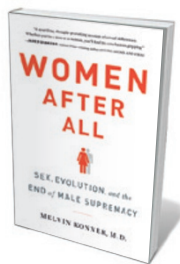
Neither Borgman nor Lohr truly grapples with the immensity of the big-data story. At its core, big data is not primarily a business or research revolution, but a social one. In the past decade, we have allowed machines to act as intermediaries in almost every aspect of our existence. When we communicate with friends, entertain ourselves, drive, exercise, go to the doctor, read a book — a computer transmitting data is there. We leave behind a vast cloud of bits and bytes.

Bruce Schneier, a security analyst known for designing the Blowfish block-cipher algorithm — a fast and flexible method of encrypting data — grasps this revolution’s true dimensions. In *Data and Goliath*, he describes how our relationships with government, corporations and each other are transformed by ordinary, once-ephemeral human interactions being stored in digital media. The seemingly meaningless, incidental bits of data that we shed are turning the concept of privacy into an archaism, despite half-hearted (and doomed) regulations to protect “personally identifiable information”. As science-fiction pioneer Isaac Asimov wrote some 30 years ago: “Things just seem secret because people don’t remember. If you can recall every remark, every comment, every stray word made to you or in your hearing and consider them all in combination, you find that everyone gives himself away in everything.”

Schneier paints a picture of the big-data revolution that is dark, but compelling; one in which the conveniences of our digitized world have devalued privacy. Interest in privacy has dropped by 50% over the past decade — at least according to Google Trends. ■

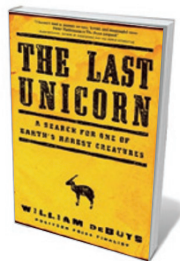
**Charles Seife** is a professor of journalism at New York University, and the author of books including *Proofiness* and *Virtual Unreality*. e-mail: cs129@nyu.edu

## Books in brief



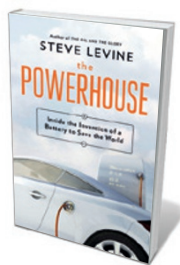
### **Women After All: Sex, Evolution, and the End of Male Supremacy** Melvin Konner W. W. NORTON (2015)

The mammalian body plan is basically female, and maleness is a syndrome. So declares anthropologist Melvin Konner in this biologically based study (although recent research points to complexities; see *Nature* **518**, 288–291; 2015). Positing that women are more altruistic and pragmatic — and so are best-equipped for the future — Konner mines evolution and anthropology to probe gender identities in the light of biology, sexual conflict across species and more. The provocative scenarios he lays out include a man-free world where women reproduce using DNA from other women’s eggs.



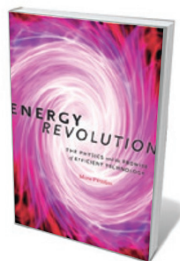
### **The Last Unicorn: A Search for One of Earth's Rarest Creatures** William deBuys LITTLE, BROWN (2015)

Discovered in 1992, the saola (*Pseudoryx nghetinhensis*) is one of the rarest large mammals, a beautiful ruminant found in the mountains between Laos and Vietnam. In 2011, nature writer William deBuys and field biologist William Robichaud set out to gauge poaching pressures on the saola. DeBuys’ account of destitute villages and endangered animals left to die in snares is a familiar narrative of conservation in poor countries. But, like Peter Matthiessen’s 1978 *The Snow Leopard* (Viking), this is less an homage to an iconic species than a meditation on our compulsion to harry and hem in the wild.



### **The Powerhouse: Inside the Invention of a Battery to Save the World** Steve LeVine VIKING (2015)

Journalist Steve LeVine’s chronicle of the race to develop a rechargeable lithium-ion electric-car battery makes for a propulsive techno-saga. The action centres on the Argonne National Laboratory outside Chicago, Illinois, where an international group led by engineer Jeff Chamberlain worked on the knotty physics. LeVine interweaves the geopolitical jostling of the US lab and others in Asia, climaxing with Argonne’s 2012 win of more than US\$120 million to build the ‘Hub’ — a powerhouse intended to create a sustainable battery industry.



### **Energy Revolution: The Physics and the Promise of Efficient Technology** Mara Prentiss HARVARD UNIVERSITY PRESS (2015)

In this crisp, evidence-based treatise, physicist Mara Prentiss makes a remarkable assertion: that solar and wind power could supply 100% of average US energy needs for the next 50 years. Prentiss argues that a transition to renewables is probable, given that energy revolutions are a historical norm. She stacks up reams of salient data, such as the fact that US energy use per capita has remained steady since 1965, thanks to increasing fuel efficiency. Although optimistic, her analyses of energy sources, combinations, conservation and storage compel.



### **House Guests, House Pests: A Natural History of Animals in the Home** Richard Jones BLOOMSBURY (2015)

Urban nature lovers relish the sight of birds or hedgehogs in their gardens. “Something odd, though, happens at the back door,” notes Richard Jones — and that is zero tolerance for wild unbidden guests, from tapestry moths to rats. Jones, a fellow of the Royal Entomological Society, is a learned guide to this alarming panoply of intruders, from the bacon beetle (*Dermestes lardarius*), a vagrant of old-fashioned larders, to the noisy edible dormouse (*Glis glis*), which can infest the attics of rural houses. **Barbara Kiser**