BOOKS & ARTS

A guide to the day of big data

Michael Nielsen enjoys a rich and stimulating collection of essays on the way in which massive computing power is changing science, from astronomy to zoology.

The Fourth Paradigm: Data-Intensive Scientific Discovery
Edited by Tony Hey, Stewart Tansley and Kristin Tolle
Published at http://research.microsoft.com/en-us/collaboration/fourthparadigm

When it came online in 1946, the US Army's giant ENIAC — Electronic Numerical Integrator and Computer — was hailed as the world's first 'electronic brain', a major step forwards in our ability to process information. It was put to work doing everything from modelling the hydrogen bomb to predicting the weather. Skip to today, and the Large Hadron Collider at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, will produce data in a single second that would take, on average, six million ENIACs to store. The Large Synoptic Survey Telescope, planned to begin operation in Chile in 2015, will produce data on a similar scale.

Hundreds of projects in fields ranging from genomics to computational linguistics to astronomy demonstrate a major shift in the scale at which scientific data are taken, and in how they are processed, shared and communicated to the world. Most significantly, there is a shift in how researchers find meaning in data, with sophisticated algorithms and statistical techniques becoming part of the standard scientific toolkit. *The Fourth Paradigm* is about this shift, how scientists are dealing with it, and some of the consequences. Its 30 chapters, written by some 70 authors, cover a wide range of aspects of data-intensive science.

The book is in four parts. The first two parts are a panorama of the new ways in which data are obtained, through new instruments and large-scale sensor networks. The fields covered range from cosmology to the environment and from healthcare to biology. Most of the chapters in these sections follow a common pattern. Each introduces a complex system of scientific interest — the human brain, the world's oceans, the global health system and so on — before supplying an explanation of how we are building an instrument or a network of sensors to map out that system comprehensively and, in some cases, to track its real-time behaviour.

We learn in one chapter, for example, about steps towards building a complete map of the human brain — the 'connectome'. Another chapter describes the Ocean Observatories Initiative, a major effort funded by the US National Science Foundation to build an enormous underwater sensor network in the northeast Pacific, off the coasts of Oregon, Washington and British Columbia. And so on, example after example.

CERN

A data-storage facility at CERN hints at the huge scale of the information revolution.

This repetition was, for me, the most enjoyable part of the book. It illuminates common questions that are being asked across these superficially very different fields: who owns the data gathered? How should their release be managed? How should they be curated? How will we preserve them for future generations? Most of all: how can we understand the data?

In parts three and four of the book, these same questions return, from the broader perspective of how the answers could and should be reflected in scientific institutions. Part three tackles infrastructure requirements, and part four looks at scholarly communication. Topics include the technical challenges of doing large-scale data analysis, such as multicore and parallel computing; workflow tools that simplify data analysis and make experiments and analysis more reproducible; and the difficult social and technical challenges of moving to a world in which large data sets are routinely published as part of the scientific process and then integrated with other data sources. The most interesting theme that emerges here is a vision of an increasingly linked web of information: all of the world's scientific knowledge as one big database.

The book has some minor shortcomings. At times, it reads too much like a brochure — perhaps inevitable, given that nearly half of the contributing authors come from Microsoft. Many of the essays assume that progress comes mostly from big grants and massive centralized programmes, an assumption not justified by the history of networked innovation. Think of the Internet, or the preprint server arXiv hosted by Cornell University in Ithaca, New York, or the gene-sequence database GenBank — each started by individuals with limited institutional support.

I also found myself wishing that the scope was broader. Science is about more than data: it is about ideas, explanations

and people. The same tools that are driving data-intensive science are also changing the nature of scientific collaboration, and these two changes are closely related. This shift in how scientists team up to create meaning is addressed in only a few chapters.

These are minor criticisms. The rise of 'big data' is one of the major scientific stories of our time, and *The Fourth Paradigm* offers a broad view that is both informative and stimulating. Better still, the book has been released under a Creative Commons licence, and is available for free on the Internet. ■

Michael Nielsen is a writer and physicist based in Toronto, Canada. His book *Reinventing Discovery*, about the impact of online tools on science, is due to be published in 2011.
e-mail: mn@michaelnielsen.org

How to get your message across

Don't Be Such A Scientist: Talking Substance in an Age of Style
by Randy Olson
Island Press: 2009. 208 pp. $19.95, £12.99

Am I Making Myself Clear? A Scientist's Guide to Talking to the Public
by Cornelia Dean
Harvard University Press: 2009. 288 pp. $19.95, £14.95, €18.00

The gulf between science and the rest of the world seems to be widening. If you think that keeping your head down, doing your research and not attempting to bridge that gap is enough, two books might convince you that science needs your voice — now.

The first is Randy Olson's *Don't Be Such A Scientist*. Olson was a tenured professor of marine biology at the University of New Hampshire in Durham before packing in his job, packing up his life and moving to Hollywood to learn how to make films. He passes on everything he's learned and saves you the trouble of the embarrassment he experienced as a scientist being cut down to size by film types.

Although the book focuses mainly on making and watching films, it gives some excellent insight into the general areas of communication in which scientists often fail.

Olson has five areas of advice that he uses as chapter headings: 'Don't be so cerebral'; 'Don't be so literal minded';' Don't be such a poor storyteller'; 'Don't be so unlikeable'; 'Be the voice of science!' He advises that scientists need to communicate in broader terms, add some humour once in a while, not shy away from speaking about things in an emotional way, tell interesting stories and be congenial.

Olson gives an excellent explanation of why scientists often have problems communicating with the public, saying that science is a process of "attempting to falsify ideas in the search for truth" and noting that "the masses thrive not on negativity and negation but on positivity and affirmation". He postulates that, when talking to a general audience, a scientist should try to suppress any urge to be negative because it comes across as arrogant and condescending, something that will often turn an audience against the speaker. This suggestion might not be welcomed by those readers who feel that scientists should never compromise.

Olson believes that science holds the fate of humanity in its hands, and if scientists are incapable of sharing their knowledge with the public then the results could be catastrophic. As more and more people make up their minds about a subject on the basis of a speaker's style, rather than the substance of what they are saying, learning how to speak about science with style is a crucial skill.

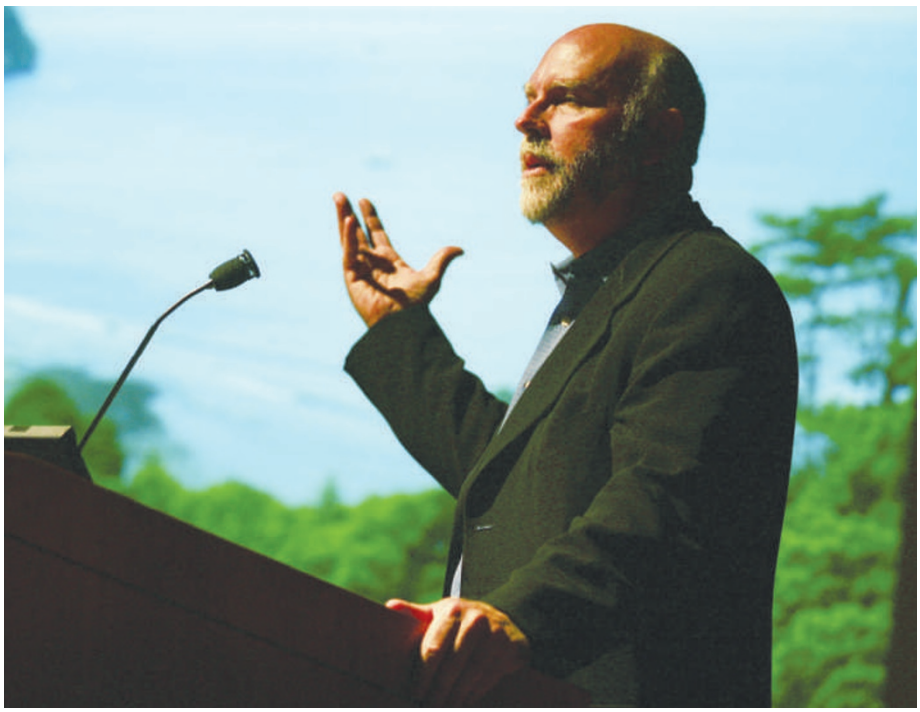
The only problem with this book is that the kind of people who need to read it are those who may be most put off by its style. Olson communicates with his audience in exactly the way he suggests — with humour, emotion and plenty of stories. Some readers may feel that it shouldn't be so much, well, fun.

If you want the facts, laid down in a simple, unfussy style, then get a copy of *Am I Making Myself Clear?* by Cornelia Dean, veteran science writer and former science editor of *The New York Times*. This book should sit on the shelf of every scientist, science communicator and university press officer. I've never read a better, more thorough guide to science communication in all its forms.

Dean's suggestions for how to be interviewed by a journalist — for print, radio and television — are spot on. From the preparation you need to do, including how to dress on TV, to always assuming everything you say is 'on the record', her book is packed full of valuable information. She also advises on producing content for the web, writing your own book and press releases, and dealing with politicians.

As Dean puts it: "We need to adopt a broader view of what it means for researchers to fulfill their obligations to society. It is not enough for them to make findings and report them in the scholarly literature. As citizens in a democracy, they must engage, and not just when their funding is at stake." ■

Gia Milinovich is a science and technology broadcaster for the BBC, Discovery Channel and Channel 4 and a new-media consultant for Hollywood films.
e-mail: giagia@gmail.com



M. NAGLE/GETTY

Genomics pioneer J. Craig Venter commands attention through passionate communication.