

The New York Times

FROM THE DIRECTORS OF  
LITTLE MISS SUNSHINE

August 11, 2012

# How Big Data Became So Big

By **STEVE LOHR**

THIS has been the crossover year for Big Data — as a concept, as a term and, yes, as a marketing tool. Big Data has sprung from the confines of technology circles into the mainstream.

First, here are a few, well, data points: Big Data was a featured topic this year at the World Economic Forum in Davos, Switzerland, with a report titled “[Big Data, Big Impact](#).” In March, the federal government announced \$200 million in research programs for Big Data computing.

Rick Smolan, creator of the “Day in the Life” photography series, has a new project in the works, called “The Human Face of Big Data.” The New York Times has adopted the term in headlines like “[The Age of Big Data](#)” and “[Big Data on Campus](#).” And a sure sign that Big Data has arrived came just last month, when it became grist for satire in the “[Dilbert](#)” comic strip by Scott Adams. “It comes from everywhere. It knows all,” one frame reads, and the next concludes that “its name is Big Data.”

The Big Data story is the making of a meme. And two vital ingredients seem to be at work here. The first is that the term itself is not too technical, yet is catchy and vaguely evocative. The second is that behind the term is an evolving set of technologies with great promise, and some pitfalls.

Big Data is a shorthand label that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases. The new data sources include Web-browsing data trails, social network communications, sensor data and surveillance data.

The combination of the data deluge and clever software algorithms opens the door to new business opportunities. Google and Facebook, for example, are Big Data companies. The Watson computer from I.B.M. that [beat human “Jeopardy” champions](#) last year was a triumph of Big Data computing. In theory, Big Data could improve decision-making in fields from business to medicine, allowing decisions to be based increasingly on data and analysis rather than intuition and experience.

“The term itself is vague, but it is getting at something that is real,” says [Jon Kleinberg, a computer scientist at Cornell University](#). “Big Data is a tagline for a process that has the potential to transform everything.”

Rising piles of data have long been a challenge. In the late 19th century, census takers struggled with how to count and categorize the rapidly growing United States population. An innovative breakthrough came in time for the 1890 census, when the population reached 63 million. The data-taming tool proved to be [machine-readable punched cards](#), invented by Herman Hollerith; these cards were the bedrock technology of the company that became I.B.M.

SO the term Big Data is a rhetorical nod to the reality that “big” is a fast-moving target when it comes to data. The year 2008, according to several computer scientists and industry executives, was when the term “Big Data” began gaining currency in tech circles. [Wired magazine published an article](#) that cogently presented the opportunities and implications of the modern data deluge.

This new style of computing, [Wired](#) declared, was the beginning of the Petabyte Age. It was an excellent magazine piece, but the “petabyte” label was too technical to be a mainstream hit — and inevitably, petabytes of data will give way to even bigger bytes: exabytes, zettabytes and yottabytes.

Many scientists and engineers at first sneered that Big Data was a marketing term. But good marketing is distilled and effective communication, a valuable skill in any field. For example, [the mathematician John McCarthy made up the term “artificial intelligence” in 1955](#), when writing a pitch for a Rockefeller Foundation grant. His deft turn of phrase was a masterstroke of aspirational marketing.

In late 2008, Big Data was embraced by a group of the nation’s leading computer science researchers, the [Computing Community Consortium](#), a collaboration of the government’s National Science Foundation and the Computing Research Association, which represents academic and corporate researchers. The computing consortium published an influential white paper, [“Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society.”](#)

Its authors were three prominent computer scientists, [Randal E. Bryant of Carnegie Mellon University](#), [Randy H. Katz of the University of California, Berkeley](#), and [Edward D. Lazowska of the University of Washington](#).

Their endorsement lent intellectual credibility to Big Data. [Rod A. Smith, an I.B.M. technical fellow and vice president for emerging Internet technologies](#), says he likes the term because it nudges people’s thinking up from the machinery of data-handling or precise measures of the volume of data.

“Big Data is really about new uses and new insights, not so much the data itself,” Mr. Smith says.

I.B.M. adopted Big Data in its marketing, especially after it resonated with customers. In 2008, Mr. Smith’s team put up a Web site to explain the Big Data theme, and the site has since been greatly expanded. In 2011, the company introduced a Twitter hashtag, #IBMbigdata. I.B.M. has a Big Data newsletter, and in January it published an e-book, “Understanding Big Data.”

Since its founding in 1976, [SAS Institute Inc.](#), the largest privately held software company in the world, has made software that sifts through databases, looking for nuggets of value. SAS, based in Cary, N.C., has seen many a marketing term in its field, including “data mining,” “business intelligence” and “data analytics.”

At first, [Jim Davis](#), chief marketing officer at SAS, viewed Big Data as part of another cycle of industry phrasemaking.

“I scoffed at it initially,” Mr. Davis recalls, noting that SAS’s big corporate customers, like banks and insurance companies, had been mining huge amounts of data for decades.

But Big Data seeks to tap all that Web data outside corporate databases as well. And as SAS’s technology has moved to exploit these Internet-era data assets, its marketing has changed, too. Last year, SAS started adopting Big Data and “Big Data analytics,” along with a term it has been using for years, “high-performance analytics.” In May, the company appointed a [vice president for Big Data](#), [Paul Kent](#).

“We had to hop on the bandwagon,” Mr. Davis says.

IT may seem marketing gold, but Big Data also carries a darker connotation, as a linguistic cousin to the likes of Big Brother, Big Oil and Big Government.

“If only inadvertently, it does have a sinister flavor to it,” says [Fred R. Shapiro](#), editor of [the Yale Book of Quotations](#).

Big Data’s enthusiasts say the rewards far outweigh the risks. Still, smart technologies that promise to observe, record and make inferences about human behavior as never before should prompt some second thoughts — both from the people building those technologies and from the people using them.

