



# < Forget YOLO: Why 'Big Data' Should Be The Word Of The Year

December 20, 2012 10:58 AM

*Copyright ©2012 NPR. For personal, noncommercial use only. See Terms of Use. For other uses, prior permission required.*

TERRY GROSS, HOST:

It gave an unheralded edge to President Obama's reelection campaign. It provokes concerns about intrusive data gathering by the government. And it's behind those irritating personalized ads that pursue us around the Web. It's called big data and it's the term that our linguist Geoff Nunberg has chosen as his 2012 Word of the Year.

GEOFF NUNBERG, BYLINE: Big data hasn't made any of the words-of-the-year lists I've seen so far. That's probably because it didn't get the wide public exposure of items like frankenstorm or fiscal cliff or YOLO. But it had a huge surge in venues like Wired and The Economist, and it was the buzz of Silicon Valley and Davos.

And if the phrase wasn't as familiar to many people as Etch A Sketch or 47 percent, big data had just as much to do with President Obama's victory as they did. Whatever it's explicitly called, the big data phenomenon has been all over the news. It's behind a lot of our concerns about intrusions on our privacy, whether from the government's anti-terrorist data sweeps or the ads that track us as we wander around the Web.

It has even turned statistics into a sexy major. So if you haven't heard the phrase yet, there's still plenty of time. It will be around a lot longer than gangnam style. It's a good name, too. I always want to pronounce it with a plosive B, the way Carl Sagan would have.

Big data, as in billions of billions of bits. That's about right for what people are calling the age of exabytes. Exabytes are what come after petabytes, which come after terabytes, which come after gigabytes. If the numbers are hard to grasp, think of it this way. If

you started stacking up those old 64k floppies until you got to 500 exabytes, you'd have a pile that stretched from here to someplace really far away.

But big data is no more exact a notion than big hair. Nothing magic happens when you get to the 18th or 19th zero. After all, digital data has been accumulating for decades in quantities that always seemed unimaginably vast at the time, whether they were followed by a K or an M or a G.

An exponential curve looks just as overwhelming wherever you get onboard. And anyway, nobody really knows how to quantify this stuff precisely. Whatever the sticklers say, data isn't a plural noun like pebbles. It's a mass noun like dust.

What's new is the way data is generated and processed. It's like dust in that regard, too. We kick up clouds of it wherever we go. Cell phones and cable boxes; Google and Amazon, Facebook and Twitter; the bar codes on milk cartons; and the RFID chip that whips you through the toll plaza - each of them captures a sliver of what we're doing, and nowadays they're all calling home.

It's only when all those little chunks are aggregated that they turn into big data, then the software called analytics can scour it for patterns. Epidemiologists watch for blips in Google queries to localize flu outbreaks. Economists use them to spot shifts in consumer confidence. Police analytics comb over crime data looking for hot zones. Security agencies comb over travel and credit card records looking for possible terrorists.

This was the year we held the first big data election. The Republicans may have had more money, but the Obama campaign had the best voter data and analytics. That gave them an edge in identifying likely supporters and finding the best ways to reach what they called low-information independents, which turned out to include running ads on Jimmy Kimmel and the rerun cable network TV Land.

And it was big data analytics that Nate Silver used to correctly predict the election outcome in all 50 states, skunking the pundits in the process. It's the amalgamation of all that personal data that makes it possible for businesses to target their customers online and tailor their sales pitches to individual consumers.

You idly click on an ad for a pair of red sneakers one morning, and they'll stalk you to the end of your days. It makes me nostalgic for the age when cyberspace promised a liberating anonymity. I think of that famous 1993 New Yorker cartoon by Peter Steiner: On the Internet, nobody knows you're a dog. Now it's more like, on the Internet, everybody knows what brand of dog food you buy.

Actually, it's more worrisome when they get your brand of dog food wrong. In some circles, big data has spawned a cult of infallibility - a vision of prediction obviating understanding and math trumping science. In a manifesto in *Wired*, Chris Anderson wrote: With enough data, the numbers speak for themselves.

The trouble is that you don't always know when to believe them. When you've got algorithms weighing hundreds of factors over a huge data set, you can't really know why they come to a particular decision or whether it really makes sense.

When I was working with systems like these some years ago at the Xerox Palo Alto Research Center, we used to talk about a 95 percent solution. So what if Amazon's algorithms decide that I'd be interested in Celine Dion's greatest hits, as long as they get 19 out of 20 recommendations right? But those odds are less reassuring when the algorithms are selecting candidates for the no-fly list.

The phrase big data may not endure. 20 years from now, when we'll probably be measuring information in humongobytes. People will be amused to recall an age when a couple of exabytes were considered big data, the way we laugh to think of a time when \$15,000 a year sounded like big money.

But 19 out of 20 is probably still going to be a good hit rate for these algorithms, and people will still feel the need to sort out the causes from the correlations. It's something you can never stop asking: What are patterns for?

GROSS: Geoff Nunberg teaches at the School of Information at the University of California Berkeley. You can download podcasts of our show on our website [freshair.npr.org](http://freshair.npr.org) and you can follow us on Twitter at [nprfreshair](https://twitter.com/nprfreshair) and on Tumblr at [nprfreshair.tumblr.com](http://nprfreshair.tumblr.com).

*Copyright © 2012 NPR. All rights reserved. No quotes from the materials contained herein may be used in any media without attribution to NPR. This transcript is provided for*

*personal, noncommercial use only, pursuant to our Terms of Use. Any other use requires NPR's prior permission. Visit our permissions page for further information.*

*NPR transcripts are created on a rush deadline by a contractor for NPR, and accuracy and availability may vary. This text may not be in its final form and may be updated or revised in the future. Please be aware that the authoritative record of NPR's programming is the audio.*

©2013 NPR