

And Data for All

Meet Vivek Kundra, the new CIO of the USA. He wants to put vast stores of government info online for anyone to use—even you.

BY NICHOLAS THOMPSON

The Library of Congress alone holds more than 300 terabytes of data—just a sliver of all federal information stores.



17.07

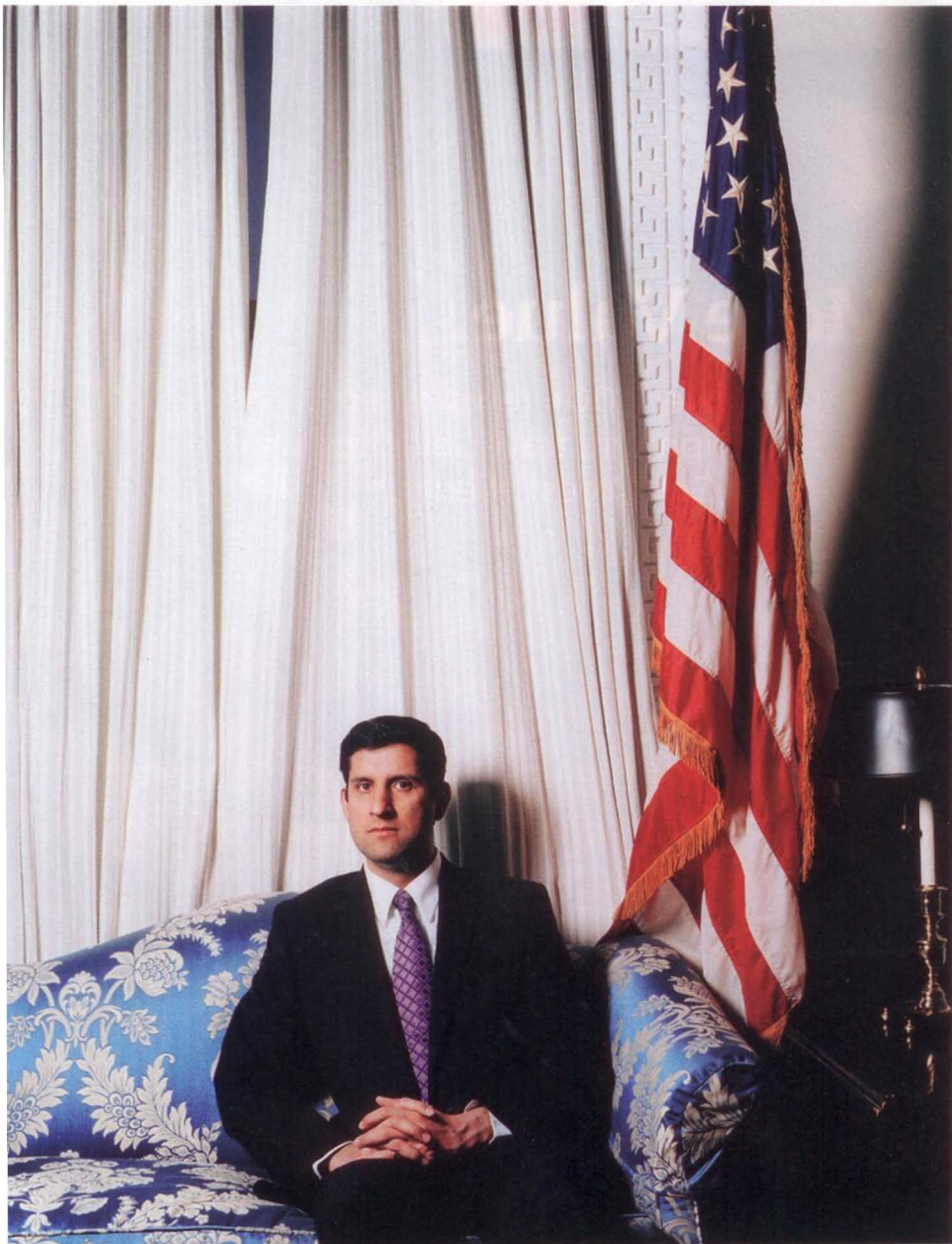
T

HE OBAMA ADMINISTRATION'S most radical idea may also be its geekiest: Make nearly every hidden government spreadsheet and buried statistic available online, all in one place. For anyone to see. Are you searching for a Food and Drug Administration report that used to be obtainable only through the Freedom of Information Act? Just a mouseclick away. Need National Institutes of Health studies and school testing scores? Click. Census data, nonclassified Defense Department specs, obscure Securities and Exchange Commission files, prison statistics? Click click. Click. Click. ¶ The man in charge is the US

government's first-ever chief information officer, Vivek Kundra. Previously CTO of the District of Columbia, Kundra, 34, knows that the move from airtight opacity to radical transparency won't be a cakewalk. Until now, the US government's default position has been: If you can't keep data secret, at least hide it on one of 24,000 federal Web sites, preferably in an incompatible or obsolete format. ¶ The goal of Kundra's new Web site, Data.gov, is to create a place where all the information is easy to find, sort, download, and manipulate. He wants to put as much data out there as possible, then sit back and let the private sector come up with great ways to use it. He envisions a future in which well-designed spreadsheets, charts, and graphs are embedded in applications for phones, Facebook, and blogs. In DC, someone combined several of the data sets released by local government—maps, liquor license info, crime statistics—into an app called Stumble Safely, which shows users the safest way to walk home when drunk. He doesn't know what people will build with all the federal data, but he's confident it will be cool. ¶ Since Barack Obama took office, WIRED has been running its own public wiki, on which scores of people have posted suggestions for how Kundra should proceed—

Kundra knows the public can create better data-driven apps than the Feds.

PHOTOGRAPH BY Ryan Pfluger



which data sets to open first, what mashups might yield interesting results, and what existing Web sites to use as models. The response suggests a real appetite for what Kundra is proposing, so we paid a visit to the White House just prior to Data.gov's launch to see how his plans are developing.

Where do you start?

One, we're going to look at which feeds are most popular and which the public are demanding. Two, we want to advance the president's agenda around health care, around energy, around education.

But won't people say you're releasing one feed because it makes Obama look good but not another that includes something embarrassing to the administration?

Well, look at health care. As the president said, it's one of the most urgent problems affecting our economic future. So it makes sense to get the most innovation in that space.

Give me an example.

There's a lot of data out there—from the National Institutes of Health, the CDC, the FDA—concerning outbreaks and pandem-

ics. And there's lots of Census Bureau data right now. For the first time, the bureau is going to be noting GPS coordinates for addresses across the country. There are privacy issues, obviously. But if you release that data at a national level, all of a sudden you've got a new layer of information that has never existed before. Imagine if you could build an iPhone app that combined the GPS info with addresses and then combined that with data about outbreaks.

You'd know precisely where outbreaks were occurring? Sort of like Google flu trends except better, because instead of search data you're using real medical data?

Exactly. And the government doesn't even have to create the applications.

What do you mean? You'll release the data and just hope people do interesting things with it?

Yes. Think about the Department of Defense. When satellite data was made available, you had this explosion in the private GPS market. Now GPS is available on your iPhone, so if you're lost you can navigate. The car rental industry uses it. Google and Facebook use it to help you get real-time information on where friends are and where the closest restaurant

is. The key is recognizing that we don't have a monopoly on good ideas and that the federal government doesn't have infinite resources. We're even thinking about running competitions for people making applications. What WIRED was able to do with that Data.gov wiki, frankly, would have cost the government a fortune and taken much longer.

Given how complicated this effort will be, are there some simple rules you're going to follow?

The core principles are using open standards, presenting raw data, and distributing it in as many formats as possible. Public policy decisions are made using the data anyway, but the raw data is important because if it is massaged too much, you can lose the big issues.

Sometimes more data confuses rather than clarifies, especially if it's raw or presented in some clumsy spreadsheet, which is typically how government data has been released in the past, if at all.

But we now have the ability to use data in ways we couldn't before, and to do it in a machine-readable way where we can not only spot trends and make intelligent decisions but make applications that create

17.07

Obama's Geek Squad

The new CIO isn't the only member of the White House tech team. Here are five other key players.

Julius Genachowski



A former executive at Barry Diller's IAC/InteractiveCorp, as well as Obama's law school classmate and tech-industry confidant, Genachowski has been nominated to chair the Federal Communications Commission. Once confirmed, he'll be in charge of developing an ambitious national broadband strategy. The Internet runs 12 times faster in Japan than in the US—at almost half the cost. Genachowski's job is to help America catch up.

Aneesh Chopra



The secretary of technology in Virginia, Chopra is Obama's nominee for CTO. He'll work with Kundra on using technology to make the administration more efficient. In Virginia he won points for helping create an online scorecard to evaluate government performance, a social network for rural medical clinics, and an open source physics textbook. And he's known for doing something rarely seen in government: punishing failure.

Susan Crawford



A onetime law professor and board member of Icanm, Crawford is now Obama's special assistant for science, technology, and innovation policy. She's also the most powerful geek close to the president. A prolific blogger before November 5, Crawford wrote passionately about NASA, net neutrality, and Neal Stephenson. Now she's helping to oversee the expenditure of stimulus package funds dedicated to broadband deployment.

Lawrence Strickling



Tapped as the new head of the National Telecommunications and Information Administration, Strickling will be in charge of managing the federal government's use of the public airwaves. Right now huge swaths of spectrum—usable for everything from Wi-Fi to cell phones to garage door openers—are locked up and hardly used by government agencies. Strickling may be able to open up some of this potential cornucopia to the public.

Alec Ross



A close associate of Genachowski during the transition, Ross now works in the State Department, looking for ways technological tools can aid diplomacy and development. Can State use mobile phones with video to monitor human rights abuses? Can social networking get Israeli and Palestinian youth to work together? Can text messaging boost HIV education in rural Africa? Ross is the person being paid to find the answers.

value and economic opportunities. The perfect example at a local level is in DC, where you can download an application that lets you know—based on where you're standing—what the closest Metro station is, when the next train is coming, and, if you like Mexican food, where the closest Mexican restaurant is. That's built on one subset of data feeds, and there are hundreds of others.

Will people be able to rate the usefulness of the data feeds?

Not only that, they'll also be able to provide feedback on quality. And one of the most important things—and this is where the WIRED community can help—is to tag the data feeds. Once you tag them, you'll be able to put them in the right context.

You mean, like tagging photos on Flickr or Google Image Labeler? So if I notice that a feed from the FAA is actually surprisingly helpful to bird-watchers, I could just note that. And then an ornithologist who's not finding what he wants through the National Park Service could see that tag?

Right.

Do you worry that all this data will come out and benefit only the few elite or tech-savvy groups that know how to use it?

Some people would say that historically there has already been asymmetrical access to the government. The key is to have debates and analysis and discussions that are fact-based. And for everyone to have access to that raw data, the raw facts. I would go back to 1776 and the model of the public square. Democratizing data enables comparative analysis of the services the government provides and the investments it makes, leading to a better government.

I can think of a hundred ways this might backfire and create terrible problems for you and Obama. What keeps you up at night?

Well, a lot of stuff keeps me up. There needs to be a balance between privacy and security on the one hand, and ensuring that we have a participatory democracy.

But if the wrong data falls into the wrong hands ... couldn't some clever hacker in Ukraine use IRS data to empty our bank accounts?

Obviously, you want to be sensitive about what you make public. We don't, for example, want to expose all of our cybersecurity information. We have to be sensible. It's a noble cause to release demographic data for research. But if you release health care data with ages and zip codes, someone may be able to triangulate and figure out who the people are. Still, the default option is to make public as much information as possible.

What about, say, nonclassified research by the Defense Advanced Research Projects Agency?

My view is that we should assess these data feeds case by case. Unclassified Darpa research sounds innocuous, but the agency may know a reason why innocuous data, combined with another feed, might be harmful. **But if the default position is open, shouldn't it be Darpa's job to argue otherwise?**

Right, that's what we're looking for.

Choosing what to open up seems like a huge task.

It is a huge task.

As CTO of Washington, you moved tens of thousands of employees from Microsoft Office to Google Apps to save money. Part of your new agenda is shifting the government to cloud computing and using free software. How will that happen?

We've got a committee working on cloud computing, and we're looking at issues like privacy, cookies, and security policies. But it makes no sense to spend billions down the line when we can get these technologies for free.

What's your bottom line. How should your agenda be judged?

Performance. By democratizing data, the American people will be able to hold their government accountable, based on evidence rather than talk. [W](#)

Senior editor **NICHOLAS THOMPSON** (nicholas_thompson@wired.com) wrote about Google in issue 17.02.