Of governments and geeks

In several countries more official data are being issued in raw form so that anybody can use them. This forces bureaucrats and creative types to interact in new ways.

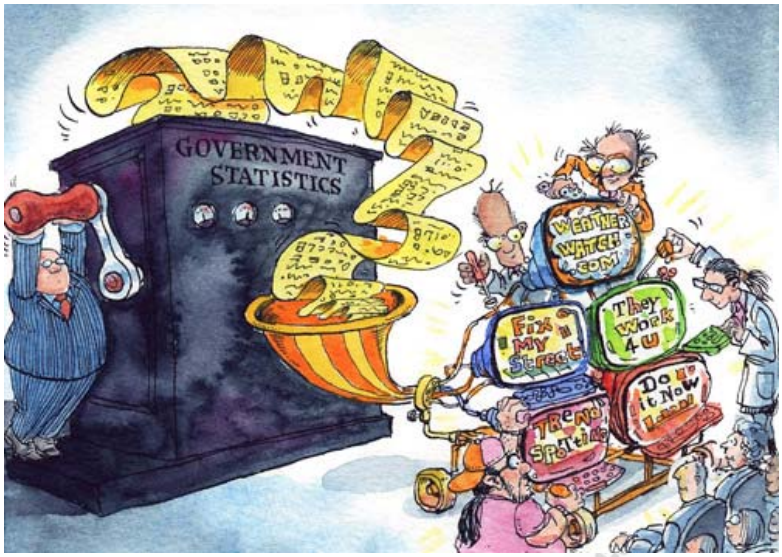

Illustration by David Simonds

YOU might think that Clay Johnson, a campaigner for transparency, would be pleased to see a ferret, with a deerstalker hat and magnifying glass, pop up on his screen. This creature is the mascot for BetaDataFerrett, an online application offered by America's Census Bureau.

In fact, Mr Johnson hates the beast. A builder of digital tools that make sense of public information, he does not need anybody to supply him with applications. All that he and his colleagues want at their Sunlight Labs—part of a non-profit group based in Washington, DC—is machine-readable data. Once he has facts that can be pulled into a computer program, he can do the ferreting for himself.

Still, Sunlight and other campaigners for better access to official information have had much to celebrate over the past year. The governments of America, Britain, Australia and New Zealand have all produced collections of machine-readable data. A British site entitled data.gov.uk was launched last month; the plan is to post a growing supply of facts that citizens or private institutions can sift through and play with as they choose.

In several countries political leaders now talk the same language as campaigners for transparent government. On his first full day in office, Barack Obama signed an open-government directive. David Cameron, the leader of Britain's Conservatives, wants to increase his country's transparency to tame the over-mighty state, for which he blames the present Labour government. In Australia Kevin Rudd's Labor Party also took power with a strong commitment to open government.

For political leaders, access to data can be a way to keep their own bureaucracies under control, and to foil foot-dragging by opponents. For example, free access to data on carbon footprints might be a way for Mr Obama to counter environmental sceptics. In his administration the task of linking information to policy—which may sometimes mean guarding data and sometimes making facts more available—has been given to two young technocrats with family roots in India. Vivek Kundra is the government's chief information officer (CIO), and Aneesh Chopra has a new job of chief technology officer (CTO).

Apart from their political uses, openly available data (about the weather, say, or from global positioning satellites) have proved valuable to many people. The British initiative, led by Sir Tim Berners-Lee, inventor of the world wide web, reflects his belief that any data can be useful.

But even as politicians start seeing the light, the pace and methods used by governments to free up facts are much influenced by independent, open-source software designers. (One reason that English-speaking governments are ahead of others is that there are a lot of activist anglophone open-source programmers.) Most of the data sets offered by governments bear the stamp “beta”, suggesting that they are open to improvement. With unusual humility, bureaucrats are borrowing jargon from open-source developers.

In the past, governments have asked large companies, like LexisNexis and Thomson Reuters, to help them handle data better. But when free, machine-readable data become available, pretty much anyone can have a go. In America, Britain and Australia, government agencies have held competitions to encourage small designers or non-profit outfits to find ingenious things to do with the stuff.

In 2008 the city of Washington, DC, a trailblazer in the field of open data, sponsored a contest called “Apps for Democracy”—with $50,000 in costs and prize money—which produced 47 applications. A competition called “MashupAustralia” has been run by the Government 2.0 Taskforce, a body set up by Mr Rudd to make administration more open. Results include a map of crime in New South Wales, and a tool for sharing data about needed road repairs called “It’s Buggered, Mate”.

All these exercises—in which anybody with a bright idea can use government data—seek to merge two cultures: the risk-averse ethos of the civil service, and the free-wheeling spirit of open-source developers, who seek continuous incremental change and see failure as a step to improvement. In a way that would baffle most old-time bureaucrats, independent developers like to collaborate over long distances and make their exchanges public.

In one culture clash, Nathan Torkington, an open-source consultant, helped New Zealand’s government assemble sets of data. After a meeting with a minister, he sent a summary of what he learned to members of his mailing list and he was gratefully accosted by subordinates of the minister who said they found out a lot about their boss.

In the world of open-source development, projects can end with a “codeathon” in which collaborators try to mesh their ideas in a burst of creative effort. This can be useful, but it differs a lot from the average bureaucrat’s working day.

Among America’s number-crunchers-in-chief, it is Mr Chopra whose office comes closest to the new culture of using data in a free, creative way. He thinks more government agencies should mimic the division of labour that now defines his own job at the top of the American administration: a CIO who guards stable information platforms and a CTO who cultivates data-handling talent in the open market.

Letting in the creative air

Gradually, government agencies are realising that contributions from small data handlers are not only tolerable, but desirable. America’s Defence Advanced Research Projects Agency is writing guidelines for contests, prizes and more open IT—procurement policies with the aim of opening up the field to smaller, more agile sellers of applications. (The Sunlight Foundation, for example, was not eligible to build recovery.gov, a data site that tracks America’s stimulus bill.)

Mr Chopra says the staff at many agencies are eager to release data, not merely to obey Mr Obama's instructions, but to gather support for their own projects. In some cases they are proud of their work and want to share it. Michael Wash, the chief information officer of America's government printing office, recalls the shock of some college students when he offered to give them the data they had been scraping off his website.

Sometimes people in government do some unlikely things—and they have a stake in making sure their work does not go to waste. (Witness the person at Australia's Department of Health and Ageing who collected a data set on the location of public toilets.) Over time, civil servants are becoming more open-minded, and developers more attuned to the needs of government. But they struggle to agree on the main question: which data have value?

Tom Steinberg, a British pioneer of data use, believes that what is valuable is what the market already pays for. He runs mySociety, a non-profit organisation founded in 2003 that builds simple web-based tools with self-descriptive names like "faxyourmp" and "fixmystreet". Location is crucial to detecting patterns in public information; a map of crimes is more valuable than a list. Britain's Ordnance Survey owns the country's geographic data and, through an arrangement known as a "trading fund" sells them to, among others, mySociety. In 2008 economic analysis commissioned by Britain's Treasury argued that the public value of the trading-fund information was greater than its revenue value to the crown; mapping data will be released free of charge from April.

But more British bulk data have been wholly or partially privatised. The Royal Mail sells postcode information, and transport timetables are sold by private rail carriers. MySociety has access to the boundary lines of voting districts through what Mr Steinberg calls "a bizarre restrictive licence". He is pleased by the data collection that has just been posted in Britain. But he notes that much of what he really needs, and now pays for, is not included.

Some special factors were at work in Britain. Dazzled, perhaps, by the magic of the Berners-Lee name, government ministers moved fast in 2009 to release whatever non-controversial information they had to hand. It is true, moreover, that Sir Tim never turns up his nose at any data, believing that even the most arcane may be handy for someone. Under his benign aegis, data.gov.uk was developed by a small group of programmers using open-source methods. The result is a geek's dream: plenty for creative types to work on, but a bit baffling to the lay person.

Some American open-data initiatives have a more user-friendly face. For example, Recovery.gov offers charts, maps and search fields. It displays some of the data on state spending that Britons lack. Local newspapers are using the site to determine how much stimulus spending has landed in their own back yards. The name "recovery" suggests an interpretation of spending that suits the president. The underlying data may be neutral, but there is always some spin in a website's presentation.

To the relief of people like Mr Johnson, the ferret-hating software developer, there is also a place where nothing but raw information is offered: data.gov, America's central collection of machine-readable data. At the same time, Mr Chopra and Mr Kundra, America's First Geeks, are considering how to respond better to requests for new data, and looking for some degree of consensus on what constitutes "high value" information. For Mr Chopra, one ambition is to find a way of linking specific budget items to actual expenditure.

But whatever governments do, the presentation of endless facts can fall flat unless there are independent developers who know what to do with them. As Mr Torkington admits, failing to grasp this point led to disappointing results in New Zealand. In his enthusiasm for technology,

he failed to think much about who would use the data he was posting, and why. A wad of facts was dumped in cyberspace, with no instructions or incentives to find good ways of using them. There they sit, unread by any machine. Even the geekiest types can be nonplussed when they are presented with data but no purpose.

Correction: We originally wrote that Sir Tim Berners-Lee invented the internet, when we meant to refer to the world wide web. Sorry. This was corrected on February 4th 2010.

Fonte: The Economist. Disponível em: <www.economist.com>. Acesso em: 11 fev. 2010.

A utilização deste artigo é exclusiva para fins educacionais