

**Australian Government** 

**Chief Scientist** 

# **PROFESSOR IAN CHUBB AC**

#### Address to the Australian Open Research Data Showcase

19 June 2015

Significance of Data Assets to the Country

Federation Ballroom North, Hyatt Hotel, Canberra I have been interested in the area of data for some time because of its potential: real potential to be realised, and potential still being realised.

Like a lot of other things that we do in science it is time for us to get our act together, to act in a more coordinated way than we have before.

Otherwise – like a lot of what we have done traditionally in our universities and science generally – when we do need to get our act together, we find it is more difficult than it should be.

I think it is also important to do so because at the end of the day, if researchers expect to get more money, if researchers expect to get more out of the National Collaborative Research Infrastructure Strategy (NCRIS) then we must demonstrate the benefits.

I think it is fair enough for the taxpayer to ask: what do we get for the investment? What are we doing? How are we positioning ourselves? How is Australia going to be better as a consequence of this investment of public money into organisations that are supposed to be working for our benefit?

And I think we have to be able to demonstrate that. We should be willing to try to demonstrate it.

I think it is important to get the message out that we do good things with that investment.

So people out there paying taxes deserve to know – if he or she wants to know – that we are doing something good with that investment. That we are making things better through making things more accessible, through making things more open, through getting our researchers to work together. That we identify areas where we need to be strong, which in a country of our size, means we can't do everything. So we need to make decisions to prioritise and invest. And to invest wisely.

#### A world awash in information

A few months ago Nature published a study on the most frequently cited papers in the Thomson Reuters collection.<sup>1</sup>

There is a corpus of some 58 million items since 1900 – and growing every day.

If that body of work were scaled to Mount Kilimanjaro, the top 100 papers would represent just one centimetre at the peak.

There are 14,500 papers – or a metre and a half – with more than 1,000 citations.

At the foothills are all the works which have never been cited, or cited only once – about half the collection.

That is some measure of the volume of information researchers have to grapple with today – as well as the amount of effort that goes largely unnoticed.

Consider the proposition for medical researchers.<sup>2</sup>

Every year, some 185,000 clinical trials take place across the world, according to the US National Institutes of Health.

One researcher cited by IBM set out to investigate the literature on a specific protein related to many cancers. He found more

<sup>&</sup>lt;sup>1</sup> Van Noorden, Maher and Nuzzo, "The top 100 papers". Nature News, October 29 2014. <u>http://www.nature.com/news/the-top-100-papers-1.16224</u>

<sup>&</sup>lt;sup>2</sup> CEDA, Australia's future workforce?, June 2015. <u>http://www.ceda.com.au/research-and-policy/policy-priorities/workforce</u>.

than 70,000 papers on this protein alone. If he could read five papers a day, he could expect to be reading for 38 years.

You get a shorter sentence for murder.

We would see something similar across every field of knowledge. Unimaginable quantities of data, enormous opportunities – and, for researchers, a daunting challenge.

## **Tapping the potential**

Late last year I convened a series of roundtables with a view to identifying the key challenges that would become our National Science and Research Priorities.

Every discipline and research community came back with the same response: we need to get better with data.

That is reflected in the final form of the priorities and practical challenges announced by the Prime Minister in May.

Data and its enabling tools are critical to every one of them.

How are we going to build smart cities without the data to understand how traffic flows?

How are we going to get mining projects off the ground without the data to maximise their efficiency?

How are we going to manage our environment – the sustainable use of freshwater or protection of our marine resources – without the data to understand and predict how these systems work, and to monitor how our actions are having an impact, or whether they are having an impact?

How are we going to increase farm yields without a wealth of knowledge from bio-sensors, gene sequencing and climate models?

How are we actually going to open up the North, in the way that is now proposed, without some understanding of the differences between growing legumes and grains down here versus up there?

Data underpins our aspirations for research – and so it underpins our aspirations for Australia.

## Accepting the challenge

But having more data is not the same as being more informed.

How much will we miss as we seek to understand our world better if our capacity to collect, store and manage data is seriously restricted?

What we make of data largely depends on what we make of ourselves: as individual researchers, and as a research enterprise.

It depends on the capacity of our people: our ICT specialists and the research workforce at large. And on data specialists – people who can do all that is necessary to enable storage, access and use of data.

It depends on the capacity of our infrastructure: in government, in the research sector and in business.

It depends on the capacity of our institutions: to enable and encourage collaboration. And it depends on the capacity of our society: to find the balance between competing interests of privacy, security and transparency.

In all of this, our people are fundamental. Our researchers must be appropriately skilled in order to optimise their use of the new data landscape.

By that, I mean all researchers, in all fields, whose work can benefit from the capacity to manage large volumes of complex data.

I don't know who wouldn't fall in that category today.

And we must have enough people with specialist ICT skills interested in working in the research sector to support them.

# A global revolution

A few months ago I saw a slide presentation prepared by some ICT professors in the United States. It was called 'Tsunami or Sea Change: Responding to the Explosion of Student Interest in Computer Science'.<sup>3</sup>

In the seven years to 2013 – 14 introductory course enrolments in ICT at MIT had nearly doubled. Demand for the major had quadrupled at Harvard and more than trebled at Stanford.

A lot of that growth had come about because people on research pathways were seeking out ICT courses and units wherever they could – understanding it to be critical to whatever projects they might pursue in the future.

<sup>&</sup>lt;sup>3</sup> Lazowska and Roberts, 'Tsunami or Sea Change? Responding to the Explosion of Student Interest in Computer Science." National Center for Women and Information Technology Summit, May 2014 <u>http://www.ncwit.org/sites/default/files/edlazowska\_ws\_lowres\_0.pdf</u>.

And they would be right.

In *this* country, there was a 55 per cent decline in the number of ICT course completions between 2003 and 2012.<sup>4</sup>

Something to ponder.

Other countries – across Asia, and in Europe, the UK, and the United States – are aware of the changing world of data. They are looking at their data assets and choosing to be leaders in research... and leaders in the global economy.

Australia wants to be at the leading edge too – to have the best infrastructure that attracts the best researchers, and generates the best scientific outcomes.

We want to be leading in order to be valuable partners for researchers from across the world.

When we think of some of the major challenges humankind has faced – and scientific advances we have achieved – over the last ten years, we see a pattern of increasing cooperation.

The Human Genome Project is one example. If we chart its development from the 1990s through to now, we can see how this important endeavour both promoted and relied on the need for researchers and others to work together to arrive at a complete sequence.

Collaboration on the Human Genome project was essentially about data. It was about amassing the collective knowledge about human genes, and sharing that knowledge – that data – with others.

That data – that formed the first complete genetic blueprint to build a human being – still diffuses into medical research labs across the world. It has an enduring impact as we continue to unlock its potential.

And so we need to develop in research the skills to work with others; a culture that encourages people to do it; and the infrastructure to support their efforts.

## A whole-of-government approach

To have that capacity we are going to have to change the way we think and the way we plan to go about things.

That means we have to take the long view.

Yes, we should be concerned that we currently have no consistent, overarching plan for our data. We need one, and it needs to sit within a broader strategy for science and research.

Yes, we should be agitating for a plan that makes better use of our e-research infrastructure and makes it easier to work out where the next big investments need to be.

But we also need to be agitating for a whole lot of other things that we don't see unless we pull back the viewfinder.

One of those things is a better approach to maths, science and ICT in schools.

The raw number of year 12 students taking mathematics is higher than ever, but the vast majority (or 64% in 2012) now take it only at the general or elementary level.<sup>5</sup>

Over the last 10 years, numbers in computing subjects at school have dropped in all states. They have more than halved in NSW and Victoria. Numbers studying mainstream computing have dropped by 70%.<sup>6</sup>

That's a problem, and it will limit our research capacity well into the future unless we address it.

Here's another one: the skills gaps in the industries where we might look to apply the fruits of our research.

A paper released by Deloitte this week suggests we will need an extra 100,000 ICT skilled people by 2020.<sup>7</sup>

We could quibble the precise number, but the momentum of change is clear: rapidly escalating needs, and largely stagnant supply.

I believe that if we are bold enough... bold enough to embrace change, bold enough to make a plan, we will be able to address some of the most pressing problems of today, using data in ways that were previously unimaginable.

And I think we are closer to that boldness than at any time in our recent history.

<sup>&</sup>lt;sup>5 5</sup> Kennedy, Lyons and Quinn, "The continuing decline of mathematics and science in Australian high schools", *Teaching Science*, Volume 60, Number 2, June 2014.

<sup>&</sup>lt;sup>6</sup> Cited in David Grover (2014). 'Reboot Teacher Training.' <u>http://theconversation.com/reboot-ict-teacher-training-to-halt-the-computing-brain-drain-25207</u>.

<sup>&</sup>lt;sup>7</sup> Deloitte Access Economics, *Australia's Digital Pulse*, June 2015.

http://www2.deloitte.com/au/en/pages/economics/articles/australias-digital-pulse.html.

The National Science and Research Priorities have been announced.

The Education Minister has secured the commitment of the State and Territories for a renewed focus on science and maths in the curriculum.

The NCRIS Review will be delivered in July.

The Cooperative Research Centre Review has been published, and the implementation phase will soon begin.

A review of Research Training is underway through the Council of Learned Academies.

Most importantly, these actions reflect a bipartisan sense that science is important.

The Prime Minister, through the Science Council, has endorsed the need for a national, whole-of-government approach.

In the coming weeks I will be consulting on the detail of the Government's response to my recommendations for Australian STEM.

So while I remain cautiously optimistic, I do sense that we have the momentum we need to make enduring changes to the way that we prepare for the future.

None of this alters the immediate proposition for a researcher at the foothills of our Kilimanjaro of information.

Perhaps it can make the summit achievable.