



Australian Government

**AUSTRALIA'S CHIEF SCIENTIST  
PROFESSOR IAN CHUBB**

**30-MINUTE KEYNOTE ADDRESS PLUS 10-MINUTE Q&A  
THIRD PLENARY OF THE RESEARCH DATA ALLIANCE**

**WEDNESDAY 26 MARCH  
9:30am - 10:15am**

**HOGAN SUITE, CROKE PARK CONFERENCE CENTRE,  
JONES' ROAD, DUBLIN, IRELAND**

Thank you Ross, for the introduction.

And thank you, Mark, for your opening remarks.

Just by looking around the room it is clear that the Research Data Alliance is a growing global community. Over 475 people registered for this three day event from all over the world, including Australia, Europe, the US, Japan, Nigeria, Sudan, India, Poland and Estonia.

That's a substantial increase from the 280 participants that attended the first plenary in March 2013 - testament to the growing importance placed on this alliance by the global community. Not just the community of people that manage the data and build the structures and systems, but also the researchers that need those structures and systems to be able to do their job properly.

If you like, that's the end game. In Australia, we often talk about "what do we do all this for? Why is it that the Australian public should invest in people to do research? Why is it an important part of the the fabric of our civilisation that we do that?" Of course one of the reasons that they do that is that it is a means to an end.

When I was a young researcher, we often used to think of ourselves as being the end game. We got our research grants, we did our research and we produced our PhD students. We did all of these things and I still think that was a very important thing to do and it was also a very important part of my life. But I guess as I've gotten older and looked at things a little differently, I see that in fact, now, I am much more comfortable with saying all of that has a purpose. The purpose is not just the doing of the research, it is the benefits that flow from the research that are the end game.

When you think about the global challenges that confront the whole of humanity, we need to actually get that end game right and we need to get the means to that end game right.

And one of the means to support high quality research that will help us get to some of the solutions that we need will be of course how we manage the immense amount of research data or data that is generated by research on a global scale these days, and it's that part of the story that I'll concentrate on in a minute.

We see representatives from governments, universities, the research sector and industry who understand the importance of working together to advance research. Research that will allow us to respond to the major issues affecting our society today.

But internationally collaborative research to respond to global challenges is not possible unless we get the fundamentals right first.

Fundamentals like sharing high quality research data.

I attended the International Conference on Research Infrastructure, or ICRI, in Copenhagen in 2012, where data and e-Infrastructure (or eResearch infrastructure as we call it in Australia) featured heavily.

And it was at about that time that the RDA was a little twinkle in the eyes of some astute forward thinkers.

Future planners who understood that we need supporting infrastructure to enable collaboration, but that underlying this infrastructure must be a will to collaborate and a will to share, and a forum to allow this all to happen.

The will to collaborate, and the will to share, is a relatively new phenomenon. It is, as was described by Mark, very often the small group or the individual working on their own that seem to be important. But things are changing and attitudes are changing.

So here we all are. Increasingly willing and able to collaborate; and increasingly willing and able to share, and a lot of that is due to the Research Data Alliance.

Research practice and endeavours have changed substantially over the last ten, even five, years.

Research is now far more global, requiring increasing cooperation between nations and individual researchers to address challenges of significance to all of us. Finding solutions to food security, a changing climate and pandemics cannot be achieved by one country alone or one researcher in one country. Global challenges will only be solved by global research endeavour.

Our increasing connectedness as scientists and researchers will also allow us to advance the boundaries of discovery beyond anything that has ever been achieved before.

It is often the case that changes in how research is conducted are a reflection of the types of collaborative technologies available at that time. Technologies that have advanced from something as mundane as the mobile handset, to the ability to remotely access the Large Hadron Collider from outside Europe, or even the big telescopes in Chile for the world's optical astronomers.

So when we think of some of the major challenges humankind has faced over the last ten years, we begin to see a pattern.

The response to the SARS virus in Asia in the early part of this century was greatly assisted by the ability of scientists in different countries to work together with government and agencies, despite being quarantined in their respective jurisdictions.

How did they do this? They transferred data to each other, using the very types of collaborative technology that are a foundation of today's research – in this case, a technology known as Access Grid, that allowed what was then referred to as videoconferencing on steroids.

Institutions participating in the response to SARS were connected via the Access Grid from countries including Taiwan, Australia, China, Korea, the United States, and this connectivity was integral to the development and deployment of solutions in the quarantine environment.

When we look at the Human Genome project and 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.

Research data is, and will be, essential to address today's questions of national and international importance. Questions that will affect each and every one of us.

Whether the global challenge is about managing our food and water assets, saving our biodiversity, predicting high impact weather disasters or addressing health issues, it's all about data.

And we need national and international, intra-discipline and inter-discipline collaboration around data to put the jigsaw together.

And I'd like to add into that mix – keeping the discipline. Just the discipline. We need to keep the discipline in what I think will be the face of pressure to do otherwise sometimes. Especially from political leadership, who in some countries expect to get access to the data without being contributors to the data pool. And I think it's very important that we try to keep as many participants as possible contributing and helping build the data pool so that it gets that widespread, worldwide recognition and significance.

And it's pleasing that you are getting on with it. Because you already know all of that. And that's exactly why you're all here.

Perhaps that's one of the benefits of being a community driven initiative – if we had waited for governments alone to get the show on the road, there would be international treaties, Memoranda of Understanding, agreements and the like and.....most likely also missed opportunities.

But perhaps I should take this opportunity to speak about the ways in which governments, or at least the Australian Government, is supporting researchers in their quest to optimise the benefits of research data.

In Australia, we have long recognised the importance of research data.

The establishment of the Australian National Data Service (ANDS), which is the Australian non-government representative in the RDA, formalises that recognition and I'll come back to ANDS a little later.

We know that data is central to all research just as you do.

We also know that we are facing a massive rise in the volume of data (some call it the data deluge) as new technologies and instrumentation revolutionise the collection and generation of even more data.

The complexity of data is also rising and requires Australia to have the cutting-edge tools needed to handle it - especially when that data is drawn and shared from different areas of research.

Our researchers must be appropriately skilled in order to optimise usage of the new data landscape.

This question of skills is one I have been talking about for some time, and I appreciate that it is an issue being faced all over the world.

In a recent speech I made to our National Press Club, I said the global market does not have just one player in it. Sometimes there seems to be a presumption in Australia that we can simply dip into the global talent pool whenever we need to fill a gap. That might have been true in the past, it might have been true that we drew many people in the past, but whether that will continue to happen as circumstances change in those other countries is too moot a point to neglect.

We have to actually have our own talent and skills security to go along with all those other securities that we talk about like energy security and food security and water security to name a few.

Skills that include the ability and know-how to collaborate effectively in today's research environment; to work to maximum benefit with others are important ones.

These skills rely not only on an ingrained desire to achieve common goals but critically on the ability to share data.

We also need to build the right environment where this important collaboration can occur.

One of the ways we're doing that in Australia is through the National Collaborative Research Infrastructure Strategy, or NCRIS, which has established the eResearch infrastructure needed to continue to support research as it becomes increasingly reliant on the increase of research data, the distribution of research data and of course access to it.

I wouldn't say that it's seamless just yet, but by consulting broadly we've developed a holistic approach aimed at providing well-integrated support to researchers.

In this way we are facilitating a better interaction between universities, public research agencies, national eResearch infrastructure initiatives and industry.

As the name suggests, the National Collaborative Research Infrastructure Strategy has enabled the Australian Government and our research sector to consider a national, collaborative approach to eResearch infrastructure investments.

One that avoids duplication, is easily accessible and is set up for researchers both now, and into the future.

I would like to see more work like this, an important part of the jigsaw puzzle, incorporated into a broader strategic approach for the development of Australian science.

I should add that the United States, the United Kingdom, and many of the European Union countries have already decided that such a strategy for science is needed. We in Australia are working towards such a strategy as I stand here.



But let me return to some of the initiatives that NCRIS supports. They include:

- high performance computing through the National Computational Infrastructure (NCI) and Pawsey Supercomputing Centre (Pawsey);
- data storage through the Research Data Storage Infrastructure (RDSI);
- data management and transformation through the Australian National Data Service (ANDS);
- data processing through the National eResearch Collaboration Tools and Resources (NeCTAR);
- data access and authentication through the Australian Access Federation (AAF); and
- data transfer through the Australian Research Education Network (AREN).

Together, these facilities and resources allow researchers to undertake complex, resource intensive research in timeframes that were previously unthinkable. Of course it is also important that they are integrated and work together, otherwise we will have the pattern of our past, which was a whole lot of programs – small and large – operating in isolation from others that should have been linked to them.

So it is a definite part of the approach that is being taken at the moment to ensure that whilst these different programs have different levels of expertise and different requirements, that nevertheless they

work together in ways that we haven't been able to accomplish in the past.

Together they allow researchers to undertake all those efforts that I referred to earlier.

So we are, for example undertaking world leading climate science, and to do that, the researchers need petascale computing capacity.

However this on its own is insufficient; they also need petascale data storage, tools to manage code and data in a suitable workflow environment as well as bandwidth to connect researchers.

In Australia, we've developed the infrastructure that enables all of that.

Our climate scientists (climate is a particular issue for Australia as we will be one of the countries quite adversely affected -it is said- by the changing climate. So it's of particular significance to us) now have supercomputing resources, good storage capacity, the ability to understand and manage the massive amounts of data created and to publish it; as well as the tools to help process data, manage codes and develop efficient workflow.

And it is the practical, everyday application of these resources that provides far-reaching important benefits.

For example, in February 2009, temperatures in some areas near Melbourne reached 48 degrees Celsius.

I don't know how many people in this room have lived with 48 degree Celsius in their suburb, but I can tell you it is fairly warm and it does have consequences. You mightn't hear of it in Dublin, you might have extremes at the other end of the scale

Record breaking temperatures during the day, particularly over multiple days and with limited respite at night, can cause severe problems for critical infrastructure.

Prolonged temperatures at that level cause damage to transport infrastructure. I'm sure you can imagine the massive disruptions that brings, since so much of Australia's freight is carried from one end of the country – north to south, east to west – on our roads and our rail systems.

The cost of power outages, transport disruptions and other consequences of the 2009 heatwave has been estimated at \$800 million. Not to mention the other disastrous impacts it had on the Australian community<sup>1</sup>.

Bushfires all over the place, hundreds of bushfires burning at one time, many of them out of control, people being burnt to death, infrastructure being destroyed – massive impact that was followed a little later by huge floods in some of the same sorts of areas. We expect that the frequency and duration of this kind of heatwave will continue to increase. We also expect that extreme weather events in Australia will increase.

Imagine a system with the computing capacity, data storage capacity and associated technology that would enable a reasonably accurate forecast of which specific infrastructure was going to buckle, fail or collapse during events such as this it would be good. One that assists planning for these events and how to manage them would be good.

---

<sup>1</sup> Queensland University of Technology 2010, *Impacts and adaptation responses of infrastructure and communities to heatwaves: the southern Australian experience of 2009*, report for the National Climate Change Adpatation Research Facility, Gold Coast, Australia.

At the National Computational Infrastructure (NCI) facility the ability to develop forecasts of this type are now being considered.

This example demonstrates why it is important to have timely and open access to data for us and for others in the world.

In Europe, in the United Kingdom, in the United States – measures have been undertaken to promote open access to research publications and open access to data.

Australia is also moving in the same direction.

In 2012 and 2013 the two principal research funding bodies in Australia, the National Health and Medical Research Council (or NHMRC) and the Australian Research Council (or ARC) made announcements regarding access policies for the research supported by each respective agency.

Both agencies require that publications arising from supported research projects must be deposited in an online repository within a 12 month period from the date of publication (with the exception of projects funded before the new rules came into effect).

Since that time, the ARC has developed new funding rules for the National Competitive Grants Program that require researchers to outline how they plan to manage research data arising from ARC-funded research and applications must include details of research data management plans.

Whilst the ARC is not mandating open data, the revised wording encourages researchers to consider the ways in which they can best manage, store, disseminate and re-use data generated through ARC-funded research.

This is good news because these types of policies help to maximise the benefits of research.

And, what's great is that Australia's infrastructure and capabilities are already set up to implement this new policy. Thanks to the work of the Australian National Data Service (ANDS), researchers are well equipped to respond positively to new funding application guidelines.

ANDS itself funds projects that make data management possible, provide training on data management policy and planning, and collect and share expertise on current practices.

It provides guides for researchers and institutions on how to manage data in order to get the most out of research, and also how to comply with grant funding obligations.

This knowledge and support is backed up by eResearch investments giving researchers access to the infrastructure they need to store, manage, use and re-use data.

This includes online repositories to upload and facilitate access to data, national storage infrastructure, to securely store and access research datasets of lasting value and importance and the tools needed to move large data sets across the country.

This combination of infrastructure and resources provides Australian researchers with opportunities for improved collaboration – both nationally and internationally.

In November last year, Australia hosted the *Third European Union - Australia Workshop on Research Infrastructure* to facilitate increased cooperation in the management and use of research infrastructure both in Australia and across the EU.

Researchers and facility operators from Australia and across the EU discussed challenges and ways to optimise the benefits of research infrastructure in areas like healthy ageing, sustainable cities and clean energy. The workshop also discussed how to improve industry linkages.

An additional day was dedicated to a *New Partnerships in Big and Complex Data Workshop*, to consider the best ways of dealing with the increasing volumes of data, and the challenges ahead for data sharing and interoperability.

It was clear from each of the workshops that the Research Data Alliance is the right mechanism to continue much of this important work.

I know that there are workshop participants here today from the EU and from Australia who have gathered with the goal of building on that cooperation and want to see the progress from those workshops continue.

And it is important work like this continues.

Let me give you an example. We often talk in Australia of being a food bowl for other parts of the world.

I am aware that there is an RDA group discussing how to gather, store, share and shift large sets of data on agriculture, in particular wheat, in order to continue collaboration on food security.

With wheat (and rice and maize) supplying more than 60 per cent of the globe's carbohydrate demand, it's on data and science that the world will now depend, to ensure it can feed a projected figure of nine billion people by 2050. Especially when already one billion people in our world go to sleep hungry every night.

From an economic perspective, Australia's wheat industry has become one of the major global wheat exporters along with the US and France.

In Australia's wheat belt, which spans across 4000km, 38,000 wheat farmers produce 38 million tonnes per year which is valued at AU\$8.5 billion, or 5 per cent of the world wheat production and 14 per cent of the world wheat trade<sup>2</sup>.

Yet our rainfall patterns are changing, our climate is shifting. Our rain is moving away from where we have traditionally grown wheat, to areas where we have never thought of growing wheat. So how we accommodate that, how we still produce food and fibre in quantities that are necessary even to meet part of our frequently articulated objective of being a food bowl, is an issue for science. And it would generate data, and the data would need to be shared.

So from my perspective as a consumer, or from my perspective as a scientist, or even from my perspective as an Australian, coordinating worldwide research efforts in the fields of wheat genetics, genomics, physiology, breeding and agronomy makes perfect sense. And we need to share it.

As a nation, Australia wants to be at the leading edge – to have the best infrastructure that attracts the best researchers and generates the best scientific outcomes. But we also want to be leading in order to attract partnerships with researchers from across the world to tackle new (and in some cases, old) complex problems of national and international significance.

To work together to address some of the most pressing problems of today, in ways that were previously unimaginable.

---

<sup>2</sup> Barlow, S., *Science for our Daily Bread*, The Curious Country, pg 58, ANU E Press, 2013.

Thanks to initiatives like the Research Data Alliance, we no longer have to only imagine what could be achieved when the best minds from around the world collaborate to solve one of those big and complex problems.

And that's why the Research Data Alliance is so important to Australia, that is why we're here. I believe that you think it is important to you too – because you are here. I think we cover enough of the world for me to be able to presume to say that this is also particularly important to the world.

If we are going to solve those problems, if we are going to address those problems, if we are even going to be able to adapt or mitigate those problems, then we are going to have to work together to do it.

We have got to put aside the historical way that we went about things – locking it up, sticking it under arcane IP rules, tying people in knots when they try to collaborate. We have got to get out of all of that and you (collectively) have got to work with us to make sure that the old ways of doing things are not the ways of the future.

I am happy to be participating in this event where jurisdictional, geographic and disciplinary barriers have no place, and happy to learn more about the global partnerships being forged and the contributions to scientific endeavours that will be accelerated through collaborating and sharing your research data. By then, we will have distinguished means from ends, and the end game is certainly worth playing for and the means are critically important for us to get there.

Thank you.