Measuring excellence in research and research training

Proceedings of a conference held at the Shine Dome, Canberra, on 22 June 2004

Australian Government

Department of Education, Science and Training

The conference, *Measuring excellence in research and research training*, was organised by the National Academies Forum and was held at the Shine Dome in Canberra on 22 June 2004.

The published proceedings, sponsored by the Australian Government Department of Education, Science and Training, are also available online at **www.science.org.au/researchexcellence**.

GPO Box 783
Canberra ACT 2601
www.science.org.au

ISBN 0 85847 223 6

Minister's Foreword

The pursuit of excellence has long characterised the Australian research community—excellence in research itself, and excellence in the training of the researchers of the future.

But how do we measure excellence or gauge the success of the pursuit?

The National Academies Forum (NAF) has contributed greatly to our understanding of the ways in which we might assess the quality and impact of research and research training.

These Proceedings, from the NAF symposium on the subject held in June 2004, will continue the debate.

In May 2004, the Australian Government announced that a Quality Framework for publicly funded research would be developed. This would measure the quality and social impact of research conducted in universities and publicly funded research agencies.

The aim is to create a Framework that will help the Government build on the already outstanding achievements of our world-class innovation system.

I have established an Expert Advisory Group to advise me on the development of the Research Quality Framework. Over the coming months, the Department of Education, Science and Training will also consult with stakeholders.

We are in the fortunate position of being able to learn from assessments that have already taken place in Australia and abroad. There is much we can learn from this work and much of the existing knowledge is brought together in these Proceedings. Reaching a consensus on the way forward is the next step.

These Proceedings will be an invaluable resource as we take that step and I commend them to everyone interested in the future of Australian research and research training.

The Hon Dr Brendan Nelson MP
Minister for Education, Science and Training

Contents

Conference welcome
Dr Jim Peacock **1**

Measuring excellence: A Chief Scientist perspective
Dr Robin Batterham **3**

Excellence in investigator-driven research
Professor Robert Graham **10**

Excellence in the humanities, creative arts and media
Professor Iain McCalman **20**

Excellence in mission-directed research
Dr Michael Barber **26**

Excellence in industry-funded research
Dr Bob Watts **36**

Excellence in social sciences in the context of impact on public policy
Dr Valerie Braithwaite **44**

Assessing excellence in research in terms of private and public benefits
Dr Geoffrey Vaughan **51**

Changing research practices in the digital information and communication environment
Mr Colin Steele **61**

Assessment criteria for excellence in research and research training

Address to breakout groups by Dr Phil McFadden **74**

Reports from breakout groups **76**

Open forum discussion **85**

Conference welcome

Dr Jim Peacock, FAA, FRS, FTSE
President, Australian Academy of Science

The National Academies Forum is a gathering of the four learned academies in Australia – the Australian Academy of Science, the Australian Academy of Technological Sciences and Engineering, the Australian Academy of the Humanities and the Academy of the Social Sciences in Australia. The National Academies Forum held a small workshop in March of this year to scope this public symposium on the topic of measuring excellence in research and research training. We are delighted that the symposium has attracted some 150 participants including scientists, scholars, administrators, research managers and policy makers from universities, research organisations and government agencies.

A number of countries, as you know, have put various schemes to measure excellence in research in place, and in some of them at least the institutions being measured, as well as collaborating with the measurement process, have also used some strategies to present the best phenotype for the measurement. We don't particularly want games-playing to affect what happens in Australia; if we have got a good, open, light-touch system it gives us the best opportunity to get things right.

Importantly, today we want to talk about research *and* research training, and it is also important to think about the recognition of emerging excellence as well as extant excellence. So I think we need to think of ways of looking at the vitality, the flexibility, of institutions and groups in research and research training.

So how do you do this? Do you make an assessment on an institution, like a university, say? Do you assess departments and then have some form of aggregation to give an institutional score? Do you assess disciplinary teams? Or do you use different assessments for multidisciplinary teams? Where does peer assessment feature? I am sure we all want to see that, and peer assessment will be very important. But we don't have that many peers in Australia, and that presents a problem in itself – or a challenge, anyway.

I think the main thing is that today we, the stakeholders, if you like, in this process, have a chance to discuss these issues together. We are not likely to come up with a single solution today, and there probably shouldn't be – nor could there be – a single solution. I think, though, we do need to recognise that ultimately the assessment of research training and research in regards to excellence is going to affect investment. So that is something that we need to hold in our minds.

This process will be a continuing process for us. We will have the opportunity, as I have said, either individually or through other arrangements that the National Academies Forum might try and put in place, to have further interactions with the people who are battling to come up with a scheme, so we have a continuing dialogue.

So what do we want out of today? We would like to know your views on what the assessment framework must achieve, for both training and research, and what it must avoid doing. I think we all have some quite strong opinions about both those things, so let's hope we are going to have a creative day.

Measuring excellence

A Chief Scientist perspective

Dr Robin Batterham, FAA, FTSE
Chief Scientist
Commonwealth of Australia

Thanks for the opportunity to kick off today's discussions. It really is an important topic.

By the way, this is a commercial which you can read whilst I am getting into the opening. I have to give it these days so that people are quite clear what my role is.

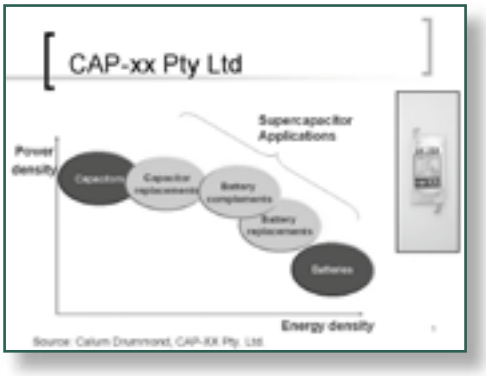
I had the benefit of quite some time last week with Dave King, the UK Chief Scientific Advisor. Dave King is quite adamant that the measuring of the excellence of research in the United Kingdom is one of the key reasons that research is now better funded in that country and taken seriously by government and wider afield – quite adamant. Indeed, coming out in *Nature* in June is a paper by David King, 'The scientific impact of nations'. It continues Bob May's earlier work showing the scientific footprint, as it is now quite developed, of different countries.

I have got the data here for Australia versus many other countries, including most of the OECD list, and it makes some interesting reading. I am not going to get into it, other than saying there are starting to be some quite well developed systems for measuring the footprint, and the excellence that sits behind that footprint, of research in countries as a whole. And I would refer you to Dave King's paper.

Let me run through a few things, because there are three points that I would like to make. They are the changing nature of science, the overall funding levels and a few comments about measuring performance.

But firstly, we don't really have a robust or a consistent way to measure the quality of research conducted in universities and publicly funded research agencies. That is what is regarded as the state of play now, and is of course the background and context for why you are here.

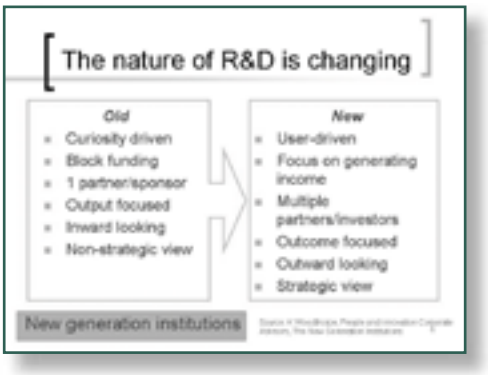
You would also all be well aware that a small amount of money has been set aside to help develop an appropriate framework.



I wanted to emphasise, particularly on the science side, the changing nature of science. I have used here an illustration from Calum Drummond, a Federation Fellow, from a presentation that he was to give to PMSEIC last week. I thought of it when I was sitting in the Institution of Electrical Engineers, in London, in this marvellous wood-panelled room with a picture of Faraday looking down at you, and you sit there and listen to whoever is presenting – it happened to be Sir David King – and doing the mental comparison, 'Is this person up to Faraday's standard or not?'

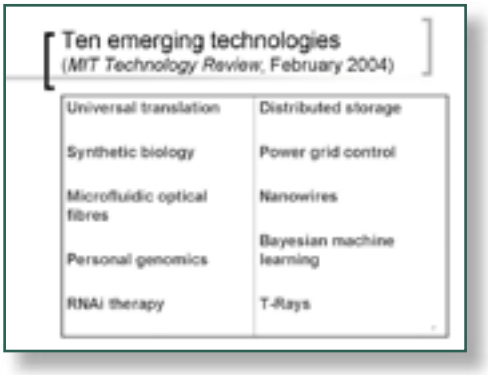
One of the intriguing things about Faraday, and ultrahigh capacity capacitors, which was the subject of Calum Drummond's talk, is that the scientific direction for how to get really high capacity capacitors was actually set by Faraday and it has taken all this time before it has been realised. And it has been realised because some great science out of CSIRO in the nanotechnology area has allowed the ultrafine carbon particles, which are at the core of these capacitors, to be produced.

The point that I am making here is that the nature of science – and wider endeavours – is changing, in that there is much more intersection across disciplines and even bridging, of course, between science, social sciences and the humanities.

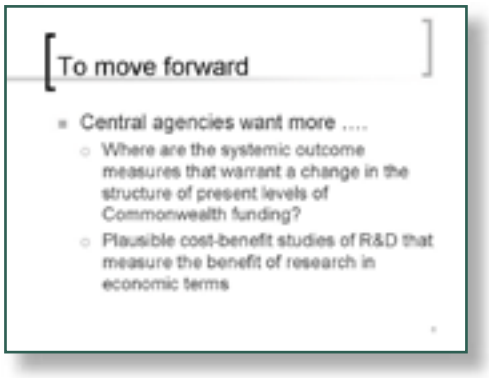


Indeed, plenty of you have heard me on this topic before, that the nature of R&D is changing. I am not arguing that all research should have a user-driven focus on generating income, et cetera. I am not arguing that at all. I am simply saying that that is becoming an emerging style of research, and it is one – an appropriate one, I hasten to add – which assessment schemes have to come to grips with.

If you look at the experience of assessment teams, it is absolutely clear that some go down to far too much detail. Gareth Roberts was commenting that that is a bit of a trap with the New Zealand system. And others are limited just to basic research, and as such they miss out on many important aspects. So we are aware of that, but the point I make over and above that is that the nature of R&D is changing and we have got to accommodate that and look at the experience of others in assessing R&D.



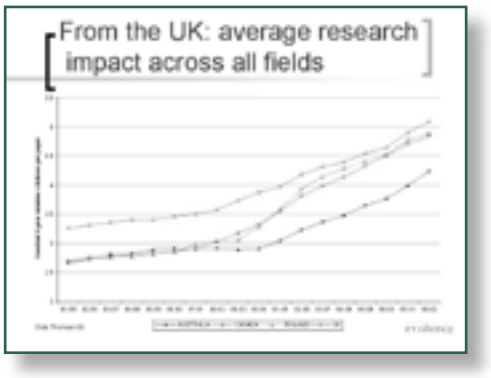
I would make two comments here. The slide shows MIT's current 10 emerging technologies. And so many of them are either cross-cutting or, essentially, disciplines that simply weren't around, even a few years ago – not all of them, but a good few of them there. So we have to have a system that copes with that.



To move forward, central agencies (by which I am talking about Treasury and Finance), want more information. I would remind you that in the UK they have now had four sets of RAEs and are looking at their fifth. I won't say anything about what it will be, but in the end of June they will be coming out with a position for their 10-year strategy for science skills and innovation. And by 'science' they define science – like Alice in Wonderland – to mean what they want it to mean. They quite explicitly, up front, talk about science, engineering, social science and the humanities. No doubt you can juggle the order if you wish, but the point is that when they talk about science skills and innovation they are talking about the full spectrum, as we are here today with the four academies. And their Treasury put science – as defined – skills and innovation up the top of the spending review: an interesting position for a country to be in.

Central agencies demand that they can have outcome measures which allow them to do meaningful comparisons across such things as health services versus R&D versus security and a string of other things. And it is not an unreasonable demand.

On the research assessment scheme, I would suggest that schemes of this nature do allow funders to assess the quality of the research. It enables you to assess your own success, it undoubtedly informs a funding model. That's the reality of life. And it is reputed to be the prompt for the enhanced performance in the UK.

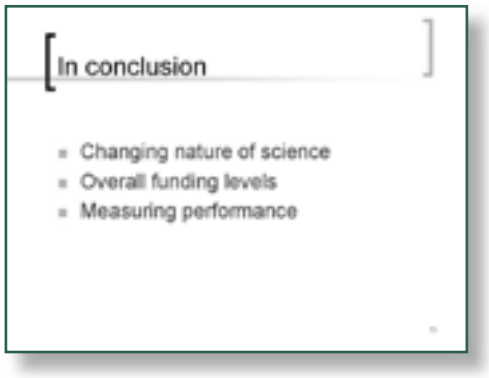


There is certainly some strong evidence, not just in Sir David King's paper coming out in June but in earlier papers on the topic. In this graph we have got Australia, Canada, Finland and the UK. The UK, which is at the top of the pack there, has shown quite some marked increase over the last 20 years. And that is what they use with their Treasury to justify more rather than less.

I commend the footprint information coming out in the *Nature* paper – and there are some earlier papers on it – because it pulls about 20-odd measures down into six general measures and then posts them up as a footprint. And it shows some improvement, I might add, over the last 20 years.

So, in tandem with what is going on today, I alert you to a few things that are also happening. On the National Research Priorities, the agencies' implementation plans are up for review at the moment; CSIRO is spending quite some time on performance measurement, and indeed has to, as part of its triennial funding arrangements and the extra money that it won there; there is a PMSEIC standing committee having a think about the broad topic of return on investment in R&D – and that is the broad return, not just the monetary return to, perhaps, individual companies – and the CCST, the Coordination Committee on Science and Technology, is also grappling with metrics in this area.

So there's the endgame, which is fairly clear. We are all, I think, interested in quality, and the question is: what value can it deliver?



In winding this up, I recite the three broad points that I wanted to make. Firstly, the nature of science as defined in that Alice in Wonderland way that the Brits define it, to be in fact the purview of the four academies, is changing. Funding levels do depend on having plausible metrics of one sort or another that you can discuss with central agencies. And can I suggest for today, in closing: don't be too abstract. By all means, have the broad discussion, but if at the end of the day there is a nice abstract position that lists the 20 things that need to be measured and the 42 reasons why none of them are perfect and why the integral of all of them is still something which can be anything, that isn't going to be too helpful.

You might consider – and this is going to be rather challenging, as my job here this morning is to be brief and to be challenging, I think you said, Mr Chairman – why you wouldn't just start with the latest version of the UK Research Assessment Exercise, and improve on that. Remember, it has had five iterations now, and an awful lot of time and effort have gone into it. So there is a pragmatic way of going, as well as the abstract one, and no doubt you will come out somewhere in between. I look forward to the results.

Discussion

Sue Richardson (President, Academy of the Social Sciences in Australia) *You put up a graph from the UK (which was also presented at a similar gathering a few weeks ago) showing how the UK performance had improved so much and attributing cause and effect. Now, to a simple social scientist that didn't look very compelling to me.*

First of all, the growth in the UK seemed to be absolutely paralleled by the growth in Australia, although the absolute levels were a bit lower. Also, the growth for Canada and for Finland looked distinctly more impressive than the growth for the UK. I wonder if you could just say a word or two about how the interpretation that this graph demonstrates the effectiveness of the UK policy has been arrived at.

Robin Batterham I am not going to try and argue how they have arrived at it. You can read the paper and look at the details, and it is worthwhile looking at. All I would make is the simple point: UK Treasury had science – as defined – skills and innovation at the head of its spending review. And I don't make any more than that. This is my point. You can be as abstract and try to be as pure as you like with it, and that's fine, but at the end of the day the prime reason why one enters such an exercise is to convince central agencies that the money they are spending is good value.

So if you can think of a better way to do that, you will have won. That's a very pragmatic answer. I hope everyone got my message.

Excellence in investigator-driven research

Professor Robert Graham, FAA, FAHA
Executive Director, Victor Chang Cardiac Research Institute

All I can say after Robin Batterham's talk is, 'Oh dear! There are lots of things I disagree with in what he said.' But anyway we will have to battle on, as you will see in my presentation.

What I chose to do today, rather than getting into specifics and examples of how you measure excellence, or what is excellence, is to explore more the notions of quality, the concepts of what constitutes excellence, and present some principles.

To begin with, let me tell you what I think investigator-driven research is, why I think it is very important and why I think we should resist what Robin is suggesting we do with the new way of funding, which I hope will become apparent. What I would like to do is to explore just four issues: what is investigator-driven research, what is excellence in investigator-driven research, how can we foster excellence in investigator-driven research, and how can we better capture the value of research?

So let me start with the first: what is investigator-driven research? Inherent in the concept of investigator-driven research is that it is a system of research funding, largely pioneered in the United States and adopted in Australia and a few European countries. It is a system not practised in Eastern Bloc countries, Japan or many European countries. In those countries the direction and the control of research are vested in a relatively few senior people, and agendas are determined in a top-down fashion.

In the US, the bulk of research funding goes to individual investigators, thus making them unbeholden to department heads, to deans or to university politics. After applying for a grant, the individual is judged in competition with other applicants by a group of peers outside of their institution, and with the awarding of a grant the investigator becomes his or her own boss, whose success or failure depends entirely on what they can accomplish.

So investigator research is curiosity-driven inquiry initiated by an individual. With respect to biomedical research, it need not be commercial or applied research, not health-outcomes research nor clinical trials research, although it can be. But it can as well just be pure basic research.

I would suggest that, much like a democracy, it is a system that has to be constantly defended but one that we are prepared to fight for – indeed, we are prepared to die for it in some instances – because we can see its profound value.

So too investigator research is not without its problems. It is a system that engenders conflict: the conflicts of freedom versus responsibility, discovery versus utility, what we are able to do versus where we need to go, strategic initiatives versus unencumbered directions, and societal responsibility versus intellectual curiosity.

And because of these conflicts, investigator-driven research, or funding for individual investigators, is continuously under threat. As Robin told you before, there is a move away from this sort of individual research into group and collective research. And, I would suggest, that is a bad thing.

For example, although Japan is to be applauded for instigating the Human Frontier Science Program, such grants are made only to groups from several countries who can devise a project advanced enough to be divided among them. The bulk of the research funding in Japan is not investigator-driven.

In Europe, the EU requires that investigators from three or more countries find a consensus project that can be parcelled up, leaving little room for a scientist to do something utterly original and entirely unpopular.

In the UK, despite what Robin says, I would suggest that in biomedical research, at least, their research has improved because of the great injection of funds by the Wellcome Trust, which was desperately needed after the Thatcher era otherwise it might have fallen completely apart, rather than because of any measures of excellence. Indeed, the Medical Research Council is increasingly consolidating grants along the lines of the EU.

In Italy, the powerful baronial organisation of research granting agencies perpetuates fragmentation and favouritism.

What about in Australia? While the abolition of block grant funding I think is to be applauded, a considerable fraction of the NHMRC budget is already obligated to program grants, in some of which directors can select investigators and projects that might not otherwise withstand peer review. Although most of these projects serve legitimate purposes, the outlay to support them becomes a fixed-entitlement expenditure.

Of greater concern to me personally is the move to direct up to one-third of the ARC budget to ‘strategic initiatives’, an issue that I tackled Vicki Sara on in this very room over a year ago. In fairness, Vicki then explained to me that that was not her idea but an initiative that came from above.

Although the recent NHMRC review, the so-called Wills report, resulted in a much-needed increase in research funding, and for that it should be applauded, I would suggest that it was predicated largely on a business platform, with heavy emphasis on research commercialisation but with little recognition or understanding of the intrinsic value, let alone the riches, of discovery – probably reflecting in part that the head of this review was a business person rather than a scientist.

How do we resolve these conflicts? Here I can offer but three principles. Like democracy, and despite its problems, investigator-driven research has served us well. Let's not throw out the baby with the bathwater. We need to defend it, and we need to uphold its value.

Secondly, I would like to suggest that scientists, let alone governments, are very bad at looking into the crystal ball and trying to determine directions and where we should be going. Strategic paths, therefore, should provide a compass, not control the journey.

And, thirdly, I think we need to get the balance right. For example, in the United States and other countries that have successful investigator-driven research programs, only 10 to 15 per cent of the total research funding is devoted to programmatic or strategic research.

I would like to digress here a moment and read you an editorial by Arthur Kornberg, Nobel Laureate for discovering DNA polymerase, and published in *Science*, 12 December 1997. He said:

> An oft-stated reason for block grants and collective efforts is the expensive equipment and resources needed to solve the problems of major diseases. A common illusion is that 'strategic' objectives are necessary to discover the cures for cancer and AIDS, and that groups of sufficient size need to be mobilised for wars and crusades against the enemies.
>
> Nothing could be more misguided. In the history of triumphs in biomedical science, such wars and crusades have invariably failed because they lacked the necessary weapons: the essential knowledge of basic life processes.
>
> Instead, some of the major advances – X-rays, penicillin, polio vaccine and genetic engineering – have come from the efforts of individual scientists to understand Nature, unrelated to any practical objective. Basic research has been the province of the individual investigator and remains the lifeline of medicine.

Let me move on to what constitutes excellence in investigator-driven research. I agree entirely with Robin Batterham that this is a very difficult issue to quantify, and I am not even going to try here. The best I can say is that when you see it, you know it, but it's damned hard to describe, let alone to quantify.

If you wanted to push me, I would suggest that, at least in biomedical research, excellence is research that provides essential knowledge into basic life processes and, further, it is research that allows us to *make the unimaginable imaginable*. And *that*, I would suggest, is its true and immediate utility.

To quote from a lovely essay given by Richard Klauser, the head of the National Cancer Institute, at the time of the Lasker Awards in 1998:

> Our interest in science is so often wrapped up in a need to justify the effort. We justify the scientific pursuit by 'the discovery', and we justify the discovery by its 'utility'. We thus assign to science its value and we too often miss its riches.

So how do we quantify excellence? I would suggest long term it is not that difficult. It is what one discovers, the impact to science and society, who one trains and the recognition by one's peers. Short term it is much more difficult to quantify. The two things we look to are citations and grant success, but they are also imperfect measures.

The impact to society can either be concrete, in terms of commercial outcomes or biotechnology, or it can – as importantly, I would suggest – be cognitive, in terms of either direct advances: new diagnostics; or indirect: a new understanding. For example, discoveries made on the anatomy of the renal collecting tubule in the desert rodent, which is able to concentrate water so effectively that it doesn't even need to drink (it just gets water from the atmosphere and from food metabolism), have led to important understandings of how our bodies regulate salt and water, which we use every day. Every clinician uses it. And I suspect that the seminal paper describing that is never cited any more. So you can see that citations would not be a help. It is something that we use all the time and is of enormous, immense importance to us in treating patients, but its discovery is something that has now been forgotten.

How can we foster excellence in investigator-driven research? To me there are just three issues that need to be considered: excellent researchers, excellent management and adequate funding.

How do you get excellent researchers? Paul Ehrlich, the father of immunology, suggested that to be a scientist you need four things (he said them, of course, in German): *Geschick, Geduld, Geld und Glück*, which roughly translate into skill, patience, money and luck. And so I would suggest it follows that if you want excellence in researchers you need to have excellent mentoring and training, you need tenacity, perseverance and courage, you need sufficient money and you need insight, creativity and critical thinking. One can certainly train people into critical thinking; insight, of course, and creativity are more innate.

For excellence in management I have turned to a model: the model of Max Perutz's Laboratory of Molecular Biology, which started in 1962. By the time Max died, in 2002, nine Nobel Prizes had been awarded to its faculty, four Orders of Merit, which is the highest order given in the United Kingdom, eight Copley Medals and an astounding 100 Fellowships of the Royal Society. So I would suggest it is a reasonable model of how you manage a research facility.

Perutz's Management Model is one that I have tried to adopt in my own institution. It is fairly straightforward: choose outstanding people and give them intellectual freedom, show genuine interest in everyone's work, give younger colleagues public credit, enlist skilled support staff, facilitate the interchange of ideas, have no secrecy – or, as we would say today, be transparent – and engender a happy environment where people's morale is kept high.

Adequate funding is a topic that for me was very difficult. What is adequate funding? Clearly, if you want to do internationally competitive research, it is expensive and you need internationally competitive funding. But I thought I would tackle this issue again more conceptually, rather than more specifically, and again read to you from Klauser's wonderful essay. He said:

> If we are to continue to support the type of beautiful science we celebrate today [and he was talking about the discovery of the cell cycle made by Paul Nurse, Lee Hartwell and Yoshio Masui] and enable the expression of the creativity of the scientists we honour today
>
> – it will take much more than well disposed appropriations committees, even more than vocal advocates;
>
> – it will take a society that truly values science and scientists, much as Renaissance Florence valued its art and its artists;
>
> – it will take a society that sees its own historical narrative as, in part, the narrative of its science.

I would suggest we have a big problem here. We have a society that sees its own historical narrative as in part the narrative of its sports. And there is nothing wrong with that. We have a society that sees its own historical narrative as in part the narrative of its brave and proud fighting men and women. But we do not yet have a society that sees, as its historical narrative, the narrative of its science.

It is a difficult issue to get around, and all I can suggest is that we need to engender an appreciation not only of the *value* but of the *riches* of science.

For example, what is the value of a gold medal? What is the value of discovering the theory of relativity? It has no intrinsic value, nor does poetry or music, yet it is very, very important to our lives and I think few of us have trouble seeing that it is important.

So we need to 'tell the stories of science'. All of us here, as scientists, if we want to get this message out, need to find ways to tell the stories of science and get people excited. The issue of gold medals, of course, is one that everyone can relate to. The fact is we know we can't run 100 metres in 10 seconds and so we appreciate somebody who can do it. We can understand poetry because it is something we can come to grips with. Most people in the everyday community can't understand science because we don't tell the stories in language that is understandable.

Here let me recite, again from Klauser's article, what I think is the way to tell the stories of science:

> The work of Nurse, Hartwell and Masui told us that the cell cycle, perhaps the most basic and the most central process of life, can be studied and dissected, and its mechanisms elucidated like the gears and springs of life's cyclic clockworks. For cancer biologists this discovery has, more than most, illuminated the black box of these diseases. Genetic changes alter the mechanisms that govern the cell cycle, and the altered cell cycle allows a cell to accept genetic damage and passes it on to its progeny – an intricate and fatal choreography of life's cellular dance gone awry.

I wish I could write and communicate as beautifully as Klauser.

Finally, let me turn to the topic of how we can better capture the value of biomedical research. I agonised over this, and certainly there are a number of suggestions I can make: develop a culture that encourages interaction between basic scientists and clinicians; reward excellence in not only basic but translational research; increase funding for research; encourage, reward and support intellectual property development; increase tax incentives for biotechnology development; provide adequate seed funding for research commercialisation – and I could go on.

But I would like to suggest that we should be thinking of this in terms that can allow us to capture the value more definitively. I would like to suggest that to truly capture the value of science we have to first understand and appreciate the *scientific enterprise*. And I think that is not done by many people, because they just aren't exposed to it and don't understand it.

What do I mean by the scientific enterprise? Again I am going to turn to Max Perutz, who elegantly said:

> To be creative in science, as in the arts, cannot be organised, despite what some may think. It arises spontaneously from individual talent. Well-run laboratories can foster it but hierarchical organisations, inflexible, bureaucratic rules and mountains of futile paperwork can kill it. Discoveries cannot be planned; they pop up like Puck (from *A Midsummer Night's Dream*) in unexpected corners.

I would suggest that if we fully understand the scientific enterprise, then we can much better capture the value of research.

So let me summarise. What I have tried to do today is to tell you what I think investigator-driven research is – namely, curiosity-driven research by an individual. I have tried to suggest that to get excellence in investigator-driven research we need research that makes the unimaginable imaginable.

How can we foster research excellence? I have tried to suggest that people, management and funding are critical. And, finally, I have tried to suggest that we can better capture the value of research if we can understand and appreciate the scientific enterprise.

I apologise for not being more specific in areas and presenting rather broad overviews, but I would like to leave you with one thought. That is: *measuring excellence in science, as in art, is more of an art than a science*.

Discussion

Tom Clark (Monash University) *Thank you for your strong and yet humble presentation, Bob. I partly focus on the paradox. You were talking about excellence in research and one of the components of excellent research was adequate funding. I wonder why we are trying to seek 'adequate' funding when the object is 'excellent' research, rather than excellent funding for adequate research, or excellent funding for excellent research – or maybe adequate funding for adequate research. It seems to me that the Australian higher education system's policy struggle within our political system in Australia is to achieve the last of those four: adequate funding for adequate research. Maybe we are talking about pipe dreams.*

Bob Graham I agree with you. I think we need adequate funding for excellent research. How we achieve that, as I suggest, is not easy. I think that conferences such as this are very beneficial from that point of view.

It does worry me a little that the reason we have conferences like this is that we are a little insecure here in Australia. It is one of our innate characteristics. In 17 years living and working in the United States I was never asked, 'Is this research internationally competitive?' whereas here every second question is, 'Is it internationally competitive?' Of course our research is internationally competitive. That doesn't mean it can't be improved, and that's why conferences like this are good.

But of course I do mean adequate funding for excellent research.

Robin Batterham (Chief Scientist, Commonwealth of Australia) *Bob, I think first of all that we are actually in furious agreement, because as I realise now – and I really can't give you*

the excuse that I have been flying for 20 hours, because everyone does that – I was talking primarily about one thing, which is the block funds and how one might move in that direction, whereas you were talking primarily about competitive grants (in the language that we use here). And I would totally agree with you, by the way.

We have a dual system here, of competitive grants and block grants. And when I suggested about a year ago that because there didn't seem to be too many quality drivers or measures on the block grants, perhaps we ought to put more effort into the competitive grants, the firing line that was lined up was pretty extensive – and Treasury noted it. The way they interpreted it (this is my interpretation of their interpretation) was that, yes, competitive grants are a pretty clever way to go, because you do have some intrinsic measures of excellence associated with them. I might add, however, that the National Science Foundation did point out, in a recent survey – actually, about three years ago – that investigator-driven research requires about 30 per cent of your time in actually working up applications. But don't take that one too seriously.

The point is that we have got a dual system, and I think for one side of it it is relatively easy to see how excellence bubbles to the surface. That is the competitive side. For the block side, which is what I was addressing, it is not so easy. And that is where I think we have to put a lot more attention in.

You just might, in your considerations, consider whether you are better off with the split that we have, of block versus competitive, or whether some other model, such in as the US, where it is far more down the investigator-driven area, makes more sense. I don't want to distract you too far by saying that if you are looking for measures, you actually do have to cover both.

Bob Graham I'm glad we are not that far apart, Robin. The problem I have with strategic initiatives is that it is suggesting that somebody up there knows where we should be going. There is some credence to that, and certainly we can identify the problems, but I would suggest that the real creativity comes from investigator-driven research and we should put most of our funding in that. If we want strategic directions to be undertaken, let's have more money but not take it away from investigator-driven research that can be put into that strategic direction.

Stephen Walker (Australian Research Council) *Bob, I was fascinated by your presentation and agree with many of the points. There was one phrase that struck me. You were talking about research, and you said 'because we all see its profound value.'*

Now, I think for this audience, for this collection of people in this workshop here today, that is not going to be a difficult thing to achieve. We can probably convince ourselves, and I think the research community in general can convince themselves, of the profound value of research. The challenge in a research assessment exercise, as Robin has alluded to, is not to

convince ourselves; it is to convince the public, it is to convince the central agencies, it is to convince the government as a whole of the profound value of research.

I think that when we are talking about research assessment, when we are talking about measures, we have to keep it in mind that we are not trying to convince ourselves, we are not trying to convince our peers; we are trying to convince people who have no background in research, who largely have no understanding of research. And the challenge here is to find the language, to find the measures. And essentially it becomes a cultural problem of how to measure research, how to describe research and how to make that case in the broader community and in the broader government.

Bob Graham I agree with you. That is why I was trying to suggest that we need to tell the stories of science, because then people can see not only the value but the riches of science. The problem is that you have to speak in language that people can understand. It is easy enough for people to understand why it is great to get a gold medal or to have an Academy Award in the picture industry, but it is much more difficult for them to understand that someone like William Bragg – without being able to stand on any other giants' shoulders – was able to use X-rays to determine structure, with what was absolute genius. It was just a wonderful discovery, yet most school kids wouldn't know who William Bragg is. We need to start at the school level with stories that they can relate to, early on, and that is the only way I know of engendering an appreciation of the riches as well as the value of science.

Frank Larkins (University of Melbourne) *Bob, I wanted to ask you whether you consider that peer review in this country is robust enough to deliver adequate funding for excellent research.*

Bob Graham We are a small society and one of the weaknesses of our peer review system is that we don't have enough people. I think of the system in the United States, for example at the NIH, where you have study sections, where you have people who have got experience in particular areas and can more robustly evaluate whether a particular project merits funding in that particular area, whereas here we have got people evaluating research in one area who have very little experience in the area of the grant. It makes it a little bit weaker, but by and large I think we do a fairly good job.

It's not a perfect system – peer review never is. I suggest, though, that it is one that we should try and work on, rather than throwing away. It may be difficult, but we have got to try our best. We may have to use more foreign reviewers. That's also difficult, because they don't understand our system, but at least they can give us an informed view of whether the science that is being presented is excellent or not.

So, it is not easy, but I would suggest that we don't throw it away because it is not easy. We have got to try and work on it and continue to laud a system that has peer review.

Jim Peacock (President, Australian Academy of Science) *I think we should remember many of the things that Bob said, but particularly, in response to Stephen's point, Bob made a strong point about understanding the narrative of science, and that we need to be able to do that so that our decision makers, as well as the general public, really understand that. I think if we achieve that, we will be a long way forward.*

Excellence in the humanities, creative arts and media

Professor Iain McCalman, FAA, FAHA
President, Australian Academy of the Humanities

What is excellent research in the humanities and arts?

Innovative or original work that moves a field of knowledge forward, breaks fresh ground, stimulates further developments. It may do this by finding new areas to explore or alternative ways of examining existing problems, for example by discovering new data, interpreting new or existing bodies of knowledge in a new way; by systematically challenging the existing canons or ideologies; by experimenting with materials and ideas in search of new formulations; by questioning methodologies and forming new hypotheses and theories, by testing and implementing models and paradigms; or by acts of creative practice or performance.

Work that makes a substantive contribution to an existing field of study or discipline. Important to stress that this can and does include scholarship which is sometimes and erroneously denigrated by comparison with what is deemed research. Scholarship, as defined in the latest British RAE as, 'the creation, development and maintenance of the intellectual infrastructure of subjects and disciplines, in forms such as dictionaries, scholarly and critical editions, catalogues and contributions to major research data-bases'.

Work that feeds the development of new technologies or the application of technologies that bring benefit to peoples.

Work that is wealth creating or that represents good value for the money invested.

Excellent research in the humanities and arts may have an immediate and recognisable use-value in terms of wealth generation or social benefit or it may be ostensibly non-utilitarian. It may be pure, or academically driven investigative research, or applied research which is policy driven. It may have an ultimate significance that lies not in its applicability to any given current set of circumstances, but in its enhancement of the cultural or spiritual dimensions of the collective life of peoples. Its significance and effects may be long-term and subtle, that is in contributing to shifts of thought and values or long-term and speculative whose results may not reach immediate fruition.

How do we achieve it?

We must identify and reward research of a good standard as well as the best possible research.

Excellent research must be tested and evaluated in the public domain, through publication or other comparable means of dissemination ranging from print, through electronic and digital forms, to performance or exhibition. It needs to be assessed both by national and international peers, according to clear, fair and agreed benchmarks.

If it includes a substantial component of practice or is embedded in creative practice, as in the visual, creative and performing arts (and similar to law, engineering and parts of medicine), it must bring enhancements of knowledge and understanding in the discipline or related disciplines and incorporate a scholarly apparatus and record of research activity that enables other researchers in the discipline or related disciplines to assess its value and significance, and its methods.

We need to foster a research culture or environment that is self-reflexive, accountable and innovative. It should encourage and allow individual scholars to evolve, formulate, publish and disseminate their research both for a national and international scholarly community of specialists and for general publics, as well as research which bridges the gap between scholarly communities and publics.

It should, where relevant, support and encourage team-based and collaborative research, interdisciplinary and multi-disciplinary as well as single discipline research, and cross-institutional as well as single institutional collaborations.

It must attract postgraduate research students and encourage their completions, and attracts and fosters the work of postdoctoral students and of early career teaching staff. Such an environment must have incentives to improve future research performance for existing staff and improve the supply and demand of researchers. Assessments of research excellence must be dynamic, rewarding past performance and prospective plans or projects. It must be flexible so as to enable new fields to develop.

How do we measure such excellence?

There are four broad methods which have been variously used around the world to measure research excellence in the humanities and creative and performing arts: peer review, such in the British RAE system: self-assessment; historical ratings and algorithmic or quantitative measures. Want to focus particularly on the last because it has been the key system in Australia and because both the British and United States higher education systems are moving more and more in that direction.

Algorithmic assessment

I am not going to waste much time on the crudely quantitative system currently employed by DEST except to say that by failing to incorporate any real component of qualitative

assessment, it is worse than useless in that it skews research practices in directions that are almost uniformly calculated to prevent a healthy and innovative research environment. It rewards mediocrity than quality. I will give a few examples from my own research oeuvre.

Quantitative indicators

One quantitative indicator commonly employed is the size of research grant income, but this is not necessarily a measure of research quality in a humanities and arts context where work is less dependent than science, engineering and technology on factors such as expensive equipment. In the arts and humanities, the number and competitive bases of research grants may be a better indicator than their size.

Another quantitative measure which has gained fairly wide acceptance is the numbers of postgraduate students attracted, and their completion rates, and the numbers of postdoctoral students attracted.

A third quantitative measure which has not been widely developed but which is being tested by various of the Research Councils in Britain and by the RAE assessors is a series of esteem indicators, such as membership of learned societies; invited national and national keynote lectures; editorial boards of scholarly journals; membership of learned boards such as the ARC, government inquiries, etc. These can be adjusted according to the particular shape of disciplines, to include performances, exhibitions, consultancies, etc.

Citation indices

This system was invented more than 30 years ago by Eugene Garfield who set up the Institute for Scientific Information based at Philadelphia in the USA, and which has published three citation indices: in science, in arts and humanities and social sciences. These are alphabetical lists of scholars whose writings are mentioned in a body of scholarly periodicals. By counting the number of times a particular researcher and their work has been mentioned in this literature you arrive at a number which is interpreted as proxy of research excellence. A series of studies, especially in librarian studies field, conducted by researchers at Loughborough University, have claimed a close correlation between such indices and other systems of assessment, including peer evaluation. These studies have argued that the correlation holds across a very wide field of subjects and disciplines, including those in the humanities, and even in fields where the specialist groupings are small, the outputs slight and the citations rare. Because the system is claimed to be relatively cheap and objective compared with peer or self assessments, it is recommended as preferential to any other.

Criticism of citation indices have come from two quarters; those who find serious fault in the crudity of existing citation methods for any field of research excellence ; and those who argue

that they have a particular problem with measuring excellence in the humanities and arts, as well as some other sub-fields in other disciplines.

Discussion

Marie Carroll (Australian National University) *I certainly agree with your sentiments that we need a combination of methods of assessment, but I was particularly interested in your advocacy of self-review. I was wondering if you had any evidence at all that that is a reliable method of measuring excellence. Do you have any knowledge of whether self-review is actually reliable?*

Iain McCalman Well, it has been tried and it is being tested, both in Canada and also in Britain. It is only reliable, I think, if you can constrain it. That's the problem. You have to verify it, and that does mean the same problem as peer review, that it uses up resources and time. It has to have very definite benchmarks, as it were, which one assesses oneself against or groups assess themselves against. It certainly wouldn't work, and I don't think would be credible, as a sole system.

What I am suggesting is that it might narrow down the very cumbersome processes so that we end up focusing more specifically; the peer review elements are subjected to more intensive analysis and in a less cumbersome manner. But I agree, I think more work needs to be done on it. Fortunately, they are testing it in various ways at the AHRB and in Canada, so I guess we just have to follow and see what they come up with.

Robin Batterham (Chief Scientist, Commonwealth of Australia) *I agree very much with what you were saying there, and I would invite you, in the discussion and the thinking, to note in the UK system how they are distinguishing between peer review and expert review. I think that when we say 'peer review', actually a lot of the time we are meaning what they call 'expert review', which is subtly different.*

Iain McCalman Thanks for that, Robin. Yes, I noticed that there is a strong emphasis now on that. In the latest RAE exercise, particularly in international reviewing, they are going to put much greater emphasis on that. International reviewers are going to spend quite a sustained amount of time in the UK; there are going to be international benchmarks, both with like countries and with adjacent countries. So yes, that distinction is important to make.

Again we have the point that Bob Graham made earlier, that this is expensive and it presents some difficulties, but I think international benchmarking is one of the ways we can offset the problem of a very small group of peers here.

Sue Richardson (President, Academy of the Social Sciences in Australia) *I would be interested in your view – and I am particularly referring now to the self-assessment – whether*

there are gender biases in these different strategies for reviewing quality. Women are well known for underselling ourselves relative to men. I don't know who gets it right, but relative to men we undersell ourselves. But there may also be gender biases in the peer review if women are doing things that are a bit different from what the men are doing – the men command the heights here. I wonder whether from your perspective you have any insights you can offer us on that.

Iain McCalman No, except that you are absolutely right. Look around! So many times I have sat in senior academic meetings and seen this sea of grey suits and masculine heads, and felt intimidated, almost, because the humanities are a notoriously feminised area. I have no doubt that there are gender biases in that. There are gender biases in the career structure of women and men at the moment, still, many of which have not been caught by this system, so you can get a problem that catches early-career people as well.

I presume that is the kind of thing we really do need to do work on, and this is where we need the social sciences to look into that. All I can say is, I suppose, that a diversity of assessment procedures may cancel out some of the biases, perhaps even by setting one against the other. But the ultimate fact is that in the system at the moment its seniority right through is heavily weighted in favour of men, and it is going to perpetuate that until that weighting alters.

Mark Finnane (Griffith University) *I think Robin Batterham, at the beginning when he put up a brief indication of what the $2.8 million from DEST was to be for, contributed to a focus so far on the assessment of research.*

Nearly $600 million a year from DEST goes into research training. The $2.8 million, I understand, is going to be to investigate the quality of research and research training. I don't know if Iain has any comments on the issue as well of assessment of the quality of research training as well – which, after all, in the Australian system is strongly dependent in terms of its outcomes on a system of expert review.

Iain McCalman That is a very good point. Funnily enough, from just looking at the literature – and this may be a problem with the literature – research training doesn't seem as controversial. There isn't as much analysis of it. And I think perhaps the problem is that we just take it for granted, that the existing system is accepted. It is simply stimulated by some galvanic action to its nerve system to make it jump further and faster, and we don't look to see if it is really working in quality terms. I suspect that we need expertise looking at that. Especially, one of the things that has struck me for a long time is the isolation of our PhD system in the humanities and social sciences, and how little we have in connection with peer groups. We have tried to do something about that in our particular centre, but something like that to be built in to the system I think would enhance quality. But we have a long way to go.

Leon Mann (Academy of the Social Sciences in Australia) *Bob Graham, I think, focused very much on individual excellence and Iain, I think, spoke most recently about 'peer' and about 'team'. That complexity of individual research and team research, where we are being encouraged to collaborate and where a lot of the work that is being done across disciplinary areas, obviously is going to require teams for creativity. It adds a level of complexity in working out what it is that is the recognition of excellence. I am wondering whether the speakers this morning have got some thoughts about how we factor in the team variable in making that kind of assessment, in particular recognising that it is somewhat harder for teams and larger aggregates to work together, although work together they must.*

Iain McCalman I don't really have any answer to that. Teamwork, collaborative work of a reasonably large-scale kind, is newish in the humanities and social sciences. It is a model that has been employed in the natural sciences for much longer, and they have particular mechanisms for assessing qualities of teams. It is a new problem for us. The only direct experience I have had is when I produced this book – *An Oxford Companion to the Romantic Age* – which in DEST points earns me virtually nothing but which took five years, has 360 contributors, 150 international contributors, and was the first *Oxford Companion* in the British area to be delegated outside of Britain. It was a massive team effort, on a very small outlay. We didn't get ARC grants for it – it is scholarship, not innovative research, it seems, even though the reviewers seem to think it has changed the paradigm.

This is part of the problem, that it is not only how you assess teams but how you assess their product, and how the people involved in this get the benefits other than me. It's only my name that appears on this book, but I have a key editorial team with three or four people who deserve the accolades much more than I do.

So there are difficulties, particularly in a culture like ours which has been used to individual books and individual researchers.

Excellence in mission-directed research

Dr Michael Barber, FAA
Executive Director, Science Planning, CSIRO Corporate Centre

There is a bit of a danger this morning that we are going to cut up the research enterprise into various silos, and I guess that is almost inevitable, but then I hope the afternoon will start to pull the various threads together. So I am going to again adopt that view. I am going to try and sketch out some of the dimensions of excellence in the topic of mission-directed research.

I guess I had three objectives in the next 15 or 20 minutes. I wanted to give you some characteristics of what we regard as mission-directed strategic research; I will then use the CSIRO's National Research Flagship Program as, if you like, an exemplar of that in the Australian context, although I will draw a little bit on international comparisons and other exemplars; and I will then spend perhaps the last part of the talk talking about what does it mean to be excellent in this sort of research, what are the possible metrics and, in particular, the role of peer review.

As I have already indicated, there is a danger in what I am saying that it will take on a sort of a life of its own, and I think one of the issues clearly before us is the complexity of the system and the fact that, in a way, for Australia we need to have excellence in a whole range of different types of research. The balance of that system is part of what we need to think about and it should not be a consequence, albeit unintended, of any assessment process to unproductively destroy the balance – though we may want to come back and debate the balance a bit.

So if I put that as a precursor, I am going to focus very much on what internationally has been called mission-directed research.

What do we mean by mission-directed research? It is really R&D directed, as the title would suggest, towards some specific application or outcome area. It is not directed towards a scientific or technical endpoint; rather, towards enabling the applications or the mission to be achieved. Its phrase probably goes back to programs like NASA's man on the moon, clearly aimed at an auditable mission for which the success measure was to land a man on the moon by the end, as you might recall the John Kennedy statement, of 1970. But the outcomes are the critical part of it.

It does not imply that you cannot do excellent and, in fact, Nobel Prize-winning science – if you put it within the science context – in mission-directed research. Teasing out those dimensions is, I think, important.

What are some of the characteristics that you might associate with mission-directed research? Again these are not by any means unique to mission-directed research, and in many ways the difference as one goes across the spectrum from the investigator-initiated research that Bob Graham talked us through to this sort of research, to the research that Bob Watts will talk about later, inspired out of industry, is really a matter of degree rather than obvious cut-offs.

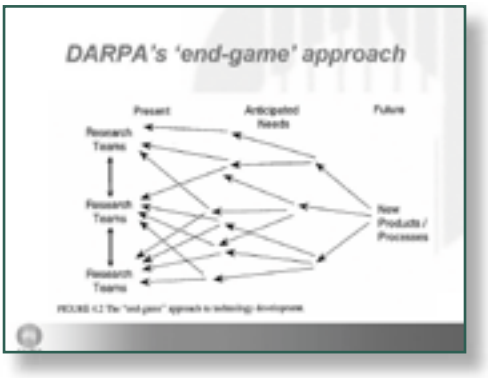
The characteristics of what I will call mission-directed strategic research are a large-scale, long-term focus on national or critical problems, usually performed by transdisciplinary teams. The objective as an outcome is to solve some problem, to deliver some outcome. They tend not to be problems which immediately define themselves as the province of chemistry, the province of biology. Generating a cotton industry for Australia, which was a mission-directed program of CSIRO's Plant Industry, didn't require just of plant scientists but also agriculturalists, across that. The disciplines that are needed are the critical part of it.

It is driven, as I have said, very much by outcomes and needs, and therefore very critically actively managed towards adoption. And I will talk a little bit more about that 'critical' thing.

The measures that you need to consider are clearly the science excellence – or perhaps more broadly today, since mission-directed research tends to involve societal impact and therefore our social scientists are becoming critical elements of the teams which address the problem – to which must be added operational excellence and in particular the outcomes and how we demonstrate outcomes. In the Australian context and again globally, these tend to be strong in the publicly funded research agencies. In fact, I would argue that without strategic mission-directed research there isn't really a logical role for a CSIRO. It is similar for ANSTO or AIMS or, in the US, the Department of Energy national laboratories, or the office of NASA's research, as examples.

There also, in a sense, are design features which I have tended to take from the following model from DARPA, the Defense Advanced Research Projects Agency of the United States. It distinguishes one of the aspects which I think we do need to distinguish.

In the Defense Advanced Research Projects Agency of the US Department of Defense, they really are mission directed. They are initiating part of the future. What are the new products and processes they need? Can they develop a detection mechanism to see through a few hundred metres of the surface of the ocean? Can you detect infrared signals? Can you specify those end products that they would want? Can you specify new materials? So it is to anticipate



the needs, come down to the present and put together the research teams that will actually deliver those outcomes and then, in particular in the mission-directed sense, manage that process to the end.

It is the complete flip of Bob Graham's description of investigator-driven research, where the ideas flow from the researchers and the research teams, are evaluated by a peer review process and lead to research outcomes which may or may not be picked up by industry or commercialisation and may or may not deliver towards those new processes. This is an attempt to take the processes and the end requirements, and drive very hard towards the end.

Peer review plays a role, and we will come back to where peer review plays a critical role. Excellence plays a critical role, but it is the other end of that dichotomy of whether it is initiated at one end under 'research teams' on the left of the slide, or whether it is initiated by the outcomes.

CSIRO's Flagship Programs, I would contend, are in fact a classical example of mission-directed research, and I want to tease out a couple of the general features of them before I turn to the question of excellence, because again these are very much part of the differentiation of the system.

They are clearly focused on major national issues. They have got 'big hairy audacious goals', to use Geoff Garrett's phrase, and they are a key driver of our alignment, in fact our response, to the National Research Priorities which we were as a Commonwealth agency required to take into account in our research planning. They have scale and impact, and will over five years consume something like 30 or 40 per cent of the appropriation of CSIRO. But they are not just about CSIRO. To build that scale we do need to reach out, to bring together 'Team Australia', to use Geoff's phrasing.

They are very much outcome-focused and outcome-directed. Their success will not just be measured by excellence in the science we do, but will be done in fact by achieving outcomes. And, critical to that, there is an active management throughout the program towards adoption.

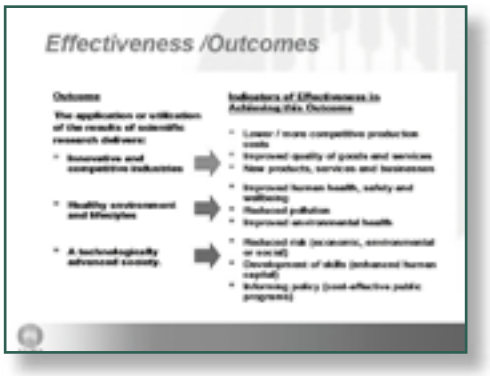
I am not going to spend too much time on the Flagships; they have been well publicised. I have just put up their missions, if you like, because I want to pull out of them a couple of key features. For example, the Flagship P-Health is to improve the health and wellbeing of Australians and save $2 billion in annual direct health costs by 2020 through prevention and early detection of chronic diseases. You can read the other ones, but for each one of them you will note that there is a quantifiable indicator of global success at some point in the future – Light Metals: to double export income and generate significant new industries, Food Futures: to transform the international competitiveness and add $3 billion annually to the Australian agrifood system.

The other three – Energy Transformed: to halve greenhouse gas emissions and double the efficiency; Water for a Health Country: to achieve a tenfold increase in the social, economic and environmental benefits of water in 2025; and the third is Wealth from Oceans.

There are two remarks I want to draw from that. One I have already alluded to. Firstly, each of those missions has a quantifiable indicator of what you might actually say is success at some time in the future. Secondly, it is not only science and technology which is going to give those outcomes, and it is not only CSIRO that can clearly deliver them. We cannot give a tenfold increase in social, economic and environmental benefits from water by 2025 with science and technology alone. Therefore, it is critical in that case that we think through the processes of adoption, the issues that need to be addressed in achieving those outcomes, because science and technology alone, and science and technology along with our social sciences that are

critical in many of these alone, are not going to achieve it. It could be regulatory, it could be tax arrangements, it could be government policies – a whole slew of issues need to be, in fact, addressed.

I am not going to drill down through the Flagships to show how these objectives are translated into detailed plans; that is a talk in its own right. But again the critical aspect is that flowing through the Flagships are these two dimensions of strategic research: a focus on the research that is going to lead to those outcomes, but secondly a critical and equal weight on what are the steps we need to take to achieve adoption so it is not something left for the future. So Tony Filmer, the director of Light Metals, whose ultimate goal is to develop a titanium industry for Australia, is already beginning to think about and talk to industry partners about what a pilot plant might look like, what a pilot plant might cost, long before the R&D that would be going into that pilot plant is in fact anywhere near fruition. The blending together of science, the research and the outcomes is critical.



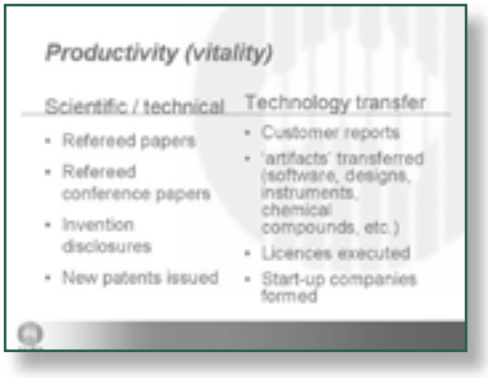
So what are some of the effectiveness/outcomes? This comes back to perhaps the most critical issue facing research assessment or research excellence, and it picks up Robin Batterham's question earlier. What are the effectiveness/outcomes?

We have in CSIRO, picking up some work we did with the Committee of International Economics, picked up what we call the 'nine shocks' – indicators of effectiveness may be lower, more competitive production costs; improved quality of goods and services; new products, services and businesses; et cetera. Again they are difficult to quantify at times, but each one is potentially possible of being quantified in some particular way. But again the link between these things and how we access the underlying science or the underlying research is a critical dimension for us to worry about.

What, then, is excellence in mission-directed research? We need measures of scientific and technical excellence, without a doubt. That picks up again many of the issues. But we also need measures of how we do the projects, technical adoption, use, economic, environmental and social impact. Putting that suite together, we need therefore to begin to tease out, I would contend, three elements which we need to build in to our assessment process.

We need to see evidence of productivity, vitality if you like, of both research results and their transfers to researchers. But also, if we are really bringing together those teams to tackle major national challenges, we need to see connectivity, both to the research community – have we brought together the necessary obvious teams to do that process, has the design of the program enabled the best researchers to in fact to be contributing to the outcome? – and critically, as I have said, in managing the adoption by the users of the research.

Finally, the impact, both in the world of science – as I have said, if we have tackled some of the nation's national challenges I believe those will in fact have pushed the paradigms of science, the limits of science, the limits of the disciplines other than science, that are necessary – but also, critically, in the intended application of the research.



So let me take each of those dimensions, if you like, and briefly – since time is of the essence – mention a few of the indicators that one might look at. In doing so, I want to pick up a couple of them.

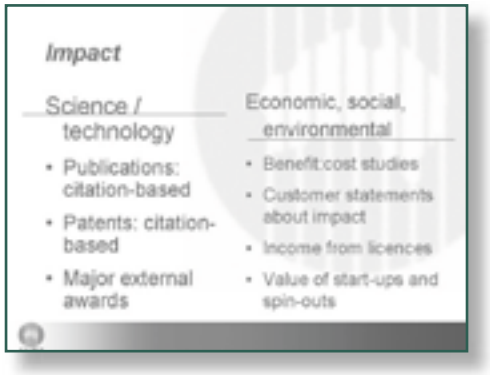
In the scientific/technical area it is relatively straightforward: clearly, where you are publishing, whether there have been invention disclosures, new patents issued. That's the scientific output, a fairly regular part of the scientific effort.

On technology transfer, again we have probably got to the point where we know something: customer reports, the artifacts transferred in software designs et cetera, the licences executed, the start-up companies formed. They give at least a sense of productivity, probably put best in the sense that if you haven't got those indicators then it is at least a valid question

to ask whether you are actually doing anything. But that is not by any means enough if we are interested in delivering the mission or the significant things.

Turning now to the research connectivity, again the connectivity to the research community is something for which we have probably got a number of metrics: students co-supervised, international visitors, the mobility, conference presentations, the engagement of professionals et cetera – all a measure or a metric that gives some sense of the engagement of the research group or the research organisation. And again customers/end-users: increasingly it is an aspect of this management to adoption that we no longer accept the linear model by which, essentially, we chuck the results over the fence and hope that industry picks them up, but to work very deliberately across organisational teams.

And again the CRCs are a good example.Their success has been to bring together the research providers and the research users to manage those activities.



And then finally, probably the most critical and perhaps in some ways the most difficult is the impact. On the left hand of the slide, under 'science/technology' without in any sense suggesting that we have got these things perfect, and echoing some of Iain McCalman's comments or concerns about the reliability or the success of citation analysis, there is a sense, at least within the sciences, that our scientific impacts are at least to some extent measurable.

A more critical issue for us is to start to think about how we get economic, social and environmental impact. There are a number of standard methods that abound, none of them perfect and all of them probably issues which could well be worked through. Here if nothing

else the engagement of our social science colleagues into the assessment of the innovation system is, I think, a critical issue. The Europeans have done some rather interesting work in this area of trying to think about the complexities of the innovation system and how information is in fact going.

You can actually do fairly standard things – benefit:cost studies, customer statements about impact, income from licences, and the value of start-ups and spin-outs. In Australia we have taken a rather simplistic view of saying, 'How many spin-offs are coming out of the University of Western Australia?' for example, or out of any university.

But I think in Australia we can see those companies. The more interesting challenge for us is whether we can generate them the next phase up – and that will be a long-term thing.

The most critical aspect about social or environmental or economic impact studies is that they are a long, long time and it is very difficult to bring together the scale on which you need to look at this research. Research for the University of Western Australia spin-off – Advanced Powder Technology Pty Ltd – goes back probably 15 or 20 years now, and its impact I suspect will be fully recognised in 10 years' time – a very long time but, I think, a natural feature of most of these impact studies.

Let me begin to draw this to an end by commenting on where the role is of peer review. We have actually said we needed all of these dimensions to be looked at. It is critical, because none of those indicators by themselves, while important, are in fact ones that stand up without interpretation. There need to be experts – I think 'expert' is probably better than 'peer' here; it includes the traditional peers but it also includes wider expert review. That judgment is necessary, judgment that is able to validate what the data says.

To me the data is critical, because in the absence of data you can at least begin to ask some very, very hard questions. Again it may be perfectly reasonable that there are no citations or there are no spin-offs out of a program that you have wanted to go, but judgment needs to be called on how you do it.

That raises the very critical issue of what we mean by peers and experts. Who are the peers, who are the experts that need to be brought to bear? I would contend again that is something that we need to give perhaps more clarity to, in the assessment process. We tend to assume that there are things called peers; we don't define very clearly the skills we require in the review committee and, since we require them to make judgment in the context of the mission, we don't define very clearly what they are going to assess. Is it a research group whose main objective is to produce basic research? That is a pretty straightforward objective. If we are looking at a mission-directed program, it is in fact important to have an appropriate focus on the mission, on the context in which we are actually doing it.

And as I perhaps indicated, in looking at mission-directed research all dimensions of excellence are equally important. So this leaves us, then, with the final part of the puzzle: when you start to put these things together across the whole of the innovation system, you have to be careful that you don't actually decide to compare an apple with an orange.

There is absolutely no doubt in my mind that by focusing on a global problem, an important national problem, you will do excellent research. You only have to look at the history of the CSIRO Division that my friend and colleague Jim Peacock led for 25 years, Plant Industry: a mission to produce value for Australia in agricultural research and in the cotton program. It has produced significant breakthroughs in many areas of plant science. But in a sense, every one of those developments had the potential to impact upon the cotton industry. You would not, for example, I suppose, Jim, have acknowledged a project in the genomics of coconuts because it had some relationship to the cotton industry. It is more focused. And so the ability to judge across these systems is a critical part of the evaluation.

It comes back again that the reviewers have to be people who have an understanding of the complexity of the science and the complexity of the research that we work in today, in part reflecting the complexity of the world.

So I guess my final message in this debate is that we have to acknowledge the complexity but, in the end, rely critically on the judgment of well-regarded, diligent peers and experts who are able to, in fact, assess on a multiplicity of dimensions.

Discussion

Doreen Clark (Australian Academy of Technological Sciences and Engineering) *Is there any facility for 'telling the story' in the system of mission-directed research?*

Michael Barber I certainly think there is a challenge to tell the story. The classic is the story of the man on the moon. I think in the Australian context there are stories which we don't actually tell: the cotton industry, on one hand; industry that is mission-directed to support the mineral industry; the values that have flowed through it – I think there are stories there that resonate very spectacularly with many of the issues, and address very much the context of those outcomes. Those outcomes have to become outcomes that resonate in a context of people's experience. I think there is a real story to be told, from industry, from CSIRO, from the wider community. CRCs, to a certain extent, tell a very good story of mission-directed research.

Chris Blackall (Australian National University) *You contrasted the mission-driven research model with Bob Graham's investigator-driven research. Clearly, in Bob's model the accountability is very much tied to the individual. If the individual succeeds, then the individual enjoys the fruit of their labour – and equally so if they screw up. In your model*

the lines of accountability are not clear, or at least they would appear to be held by the organisation, and clearly as in all organisations the rewards are distributed asymmetrically.

I was wondering how you could square up this issue of accountability and the individual contributions of researchers as reflected in citation counts and in the literature.

Michael Barber It is an important issue. To take the Flagship Programs, the accountability of the Flagship Programs, the accountability to the government for the extra money as a headline in Backing Australia's Ability, is clearly on CSIRO. It is an agency responsible for delivering them, and that impact is critical.

The question of, now, how you actually translate the values, how you get reward systems, how you get motivations of staff et cetera within that framework, is a critical issue of the human resource processes within CSIRO. Individuals will drive and deliver those Flagships, and therefore what are their internal reward systems, how do we recognise excellence et cetera?

I think for the wider science community, though, the danger is that the metrics which are actually used to achieve recognition of someone may not be appropriate. Let me pick a scientist working at the Queensland Centre for Advanced Technologies doing mission-directed research to add values for Australia's iron ore experts working for BHP Billiton on projects of importance to BHP Billiton. That person will not be publishing as extensively as someone working in investigator-driven research. I would contest that he or she is making a significant value for Australia; he or she is certainly being rewarded within CSIRO. But the potential is that they will not be perceived as being as valuable within the science story or telling as great a story, as Doreen's comment suggested, as someone working in investigator-driven research, because they are not publishing, they are not being seen so visibly in that community.

So I think those stories are important but the individual's responsibilities are different in the two situations, as part of that rich fabric of the research enterprise one needs in the 21st century.

Excellence in industry-funded research

Dr Bob Watts, FAA, FTSE
Technology Consultant, Cherrybrook, NSW

I guess it is slightly difficult for me to talk to an audience like this, because I am normally speaking in terms of assessing value to an industrial organisation rather than to what is, I think, largely academics and CSIRO. So what I will do is that as I go through I will try to draw some analogies with some of the things that I have heard this morning and, of course, that I am aware of through my contacts with universities.

The first thing I wanted to draw your attention to, in case you hadn't noticed it, is that we are existing in a world of constant uncertainty and constant change. As you will see in a few minutes, that has to be built in to any assessment of excellence.

I will just put that into context for industry. Probably most of you know that up to, say, 15 years ago, through the '80s, industry had very large, well-funded, interest-driven research labs. Famous ones were Bell Telephone Labs, with a number of Nobel Prize winners, the IBM labs and so on, and even within BHP at that time there were some 650-odd people.

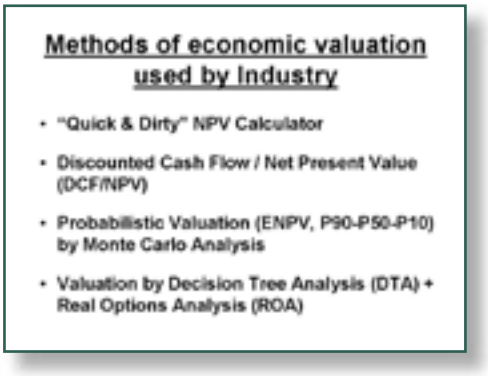
Some time in the early '90s the bean counters, as they are called, started to ask the question, 'What value is coming out of this?' And that is exactly the question, I believe, that the university community is being asked now by the public, through the politicians and of course through the Public Service: 'What value is being given to society?'

In the case of industry, the answer from the researchers was exactly the sort of arguments I have heard from other speakers this morning: 'It is self?evident that we are adding value.' That is a simple way of putting it. So the answer in industry was quite simple. They cut the funding. All those giant labs were torn to pieces. In the BHP case, in the second half of the '90s we took those 650 people down to 100 or so. And then we were asked the question, 'Well, if you want to grow, you had better show you create value.'

How do you measure value? It is quite simple in industry: it is cash. If you can put a dollar value on it, then you get somewhere. So we had to develop methods, which I will talk about in a few moments, for persuading the management that we were creating value for the company.

Scientific excellence is obviously necessary if you are going to create value through technology, but it is not sufficient to say, 'We are scientifically excellent.' You have to show what value you are creating, in terms that the management understand. And what you heard from Robin

Batterham this morning, whether you like it or not, is that he is telling you, I believe, that you are going to have to show that you are creating value for the country, using methods that the country can understand.



What we had to do was first of all to find out how the management itself valued its own activities, and then embed our technology within it. We developed within BHP – and these methods are used quite widely within industry – in our case four methods of actually putting a dollar value on what we were doing.

The first one, which we called a 'quick & dirty' net present value calculator, is essentially analogous to the way in which you calculate your mortgage repayments, or put a slab of money in and get your pension out. It is quick and dirty. We have a slab of money within BHP for curiosity-driven research – it is a pretty generous sum – and if anybody came to me and asked for funding from that they had to bring with them a very first estimate, using that first estimator: 'How much money is it going to make for the company?' Quite simple.

There are more complicated models. The next one is the discounted cash flow. Both of those models I believe have an analogy in what you are seeing at the moment from the government agencies where they are asking you to give citation indexes and so on – simple counters that they are trying to assess the value. What they have not done yet is turn it into dollars; however, dollars are embedded deeply in what you do. Whether you like it or not, every grant you get is given in dollars. So ultimately these measures will have to be mapped onto dollars.

The two methods at the bottom of the slide, on which I will not go into any detail, of course, are basically taking into account the huge uncertainties that surround research. Is it going to work? Will it not? What are the probabilities of it? And, in the industry context, will it make money? These methods were taken directly from the financial areas of the company, and essentially we have developed methods which are a generalisation of the share option process, the share option market. We treat research as an option.

I will show you that in a moment, but basically what it does is to allow you to include the huge uncertainties surrounding research, and markets, and make money from it.

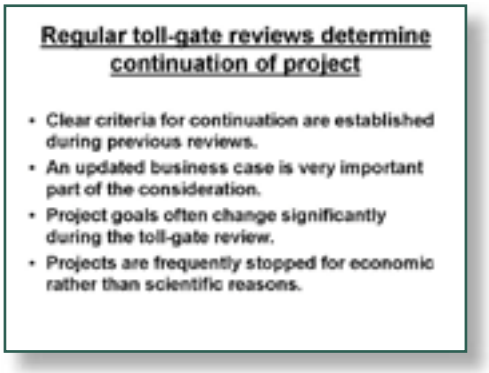
There are some pretty clear uncertainties that come about if you are trying to say how much money a new area of research is going to make. Supposing CSIRO is getting into titanium. (BHP started that as well and we already have a pilot plant in Australia.) In order to get the funding for that, which is very significant, we actually had to consider, among other things, all these distributions. We had to work out, 'What is the selling price going to be? How is it going to vary over the next 40 years?' and the production costs, construction costs and so on. Then we had to do this so-called Real Options Analysis, and by going along to the management and saying, 'By taking into account all these things, including the probability of technical success, we believe that this project can make $100 million' – or whatever the figure is – 'for the company,' we found then the management was willing to say, 'We will fund it.'

When we said less quantitative things like, 'It's obviously good to have a great titanium industry. We don't have one at the moment,' or whatever, they just shrugged.

So they fund it if you can present to them, in terms they understand, the value that in this case the company is going to make. And I repeat that I believe that what we heard from Robin Batterham this morning is basically a plea to the universities and research institutes that are government funded that they have to start thinking this way, if they want to go for more and more money. If you think about it, there is several billion a year going into universities and related institutes now – huge sums of money.

With that background, then: in industry – and I should say that the methods that are being used in BHP Billiton now have been embedded quite strongly into drug research, drug companies and so on – basically we have managed to change management's view that R&D is not a discretionary cost to be saved when times get hard but rather an investment that can add value. If you think about it, that is exactly the same type of switch that we need to bring about in government circles when they look at universities. It is not a cost to be cut, it is an investment and it can add value. But you are going to have to use their methods of thinking to show them that.

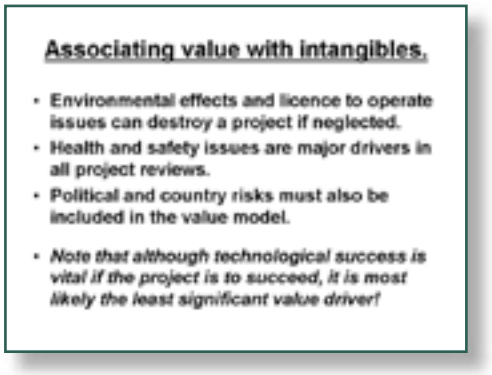
The important point in this slide that I also wanted just to mention is that the company now sees research and development as a series of 'option plays'. If you are familiar with the share market and options in that sense, you will know that you can make money by, for example, not taking up an option – in this case, by cutting a project.



All of the projects in BHP Billiton are reviewed on a continuing basis, every six months, and very interestingly the people that come and do the reviews are not the technologists, they are the management, the people who are ultimately going to use it. At each review stage there are clear criteria established, if it is going to continue beyond the next review, and at that review you ask, 'Has it met it?' Having said that, I have to point out that often at those reviews the actual direction of the research will change. In other words, the findings over the previous six months or so will affect the outcomes quite strongly. So there is very clear guidance from management, from the people who have a vested interest in the outcomes.

In Australia, within the university and national lab type context, I think Michael Barber has shown us that CSIRO is starting to move this way with its BHAGs, and in terms of the universities, the best we do at the moment is that page review you write to ARC every year to say what you have done – which is not really looked at very seriously. I think that will probably change also over the next few years. People are going to more closely look at the outcomes.

The last point on the slide is also important. In industry you will stop a project because it is not going to add value, in terms of economic value, even though the science may be doing remarkably well. And again I think what you are seeing with the National Research Priorities is some of that thinking coming into the government areas.



This slide I think has got some lessons for us all. When I gave you the uncertainties, I picked on obvious things like the price and the cost of building the plant and all the rest of it – easily quantifiable measures. Unfortunately, we are now having to grapple in industry with associating dollar values with intangibles. And much of what Bob Graham, for example, said this morning was tied up with saying, 'You can't put a dollar value on intangibles.' Well, unfortunately, at least in industry, you can put a dollar value on intangibles.

Let's look at the first one, to give you a quick example: environmental effects and the licence-to-operate issues. I am sure almost everybody in this room remembers BHP and Ok Tedi, the copper mine in New Guinea. BHP took a decision in the early '80s, after two tailings dams had fallen over, not to put up a third one. This saved many millions of dollars. Ten or fifteen years later, that came back to bite them and it bit them badly. To tell you how badly they were bitten, if you think about it, in 1996–98 the share price went from about $20 down to $12, largely driven by intangible issues like the Ok Tedi thing putting them on the television screen every night. Huge value was destroyed by something that, up front, you don't know how to put a dollar value on, except in terms of whether to spend money on a dam or not.

Health and safety issues are also now having to be quantified, as are political and country risks.

Within the Australian academic context, what we have had for many, many years is people saying, 'Interest-driven research is valuable.' I will give you two examples: 'Look what relativity did for us' – after about 70 years – and, as I heard the other day, 'Look what Tolkien did. He didn't set out to make money when he wrote *The Lord of the Ring*s, and look at the couple or three hundred millions that have been made in the past year or so for New Zealand.' Whether you like it or not, I think, if you want the money to keep increasing, you are going to have to quantify the intangibles on which we all say, 'Take it for granted.'

I will give you another example, because when Bob Graham was speaking it occurred to me. Supposing the Sydney Opera is on at the Opera House. Would you pay $5000 for a ticket? Most unlikely. Every time you buy an opera ticket you are putting value on something that I think Bob said you can't put value on – music. Yet you do. To go to some pop music, the teenager will spend twice as much as you will to go to the Opera House. So whether you like it or not, in your everyday life you put value on it, and what you are hearing from Robin Batterham and the government I think is that you are going to have to put value on it.

A couple of final thoughts for you. In industry people keep saying to me, 'Do you use outsourcing? Do you give contracts to CSIRO, or to universities, or do you do your own research?' Well, the answer to whether or not a company does that is really tied up very closely with its policies. If you have got a strong, well-funded internal R&D program, that is

invariably associated with a company policy to create value and growth through innovation. And to give you an example there with BHP Billiton, three or four years ago it didn't do that. Because after the merger of BHP and Billiton the company became so large that any regular method of growth was going to be attacked by government regulators for monopolies and so on, it was decided that a well-funded internal R&D program is necessary. The funding has tripled in the past two to three years. In other industries you will find a strong emphasis on outsourcing, but in that case the company policy is to emphasise efficient production and not to grow through innovation. So those are two thoughts.

In industry you need a balanced technology portfolio, with revolutionary technologies – for example, the titanium exercise I mentioned earlier – which are going to drastically change cost structures and introduce new products and processes; you need some evolutionary stuff which brings about significant improvements in efficiency to established methods; and you have to have some tactical technology supports for when something seriously goes wrong and has got to be fixed in a moment.

Those methods I had a few minutes ago of estimating the value of technology are all applied to these things in one form or another.

Here are a couple of comments on the roles of universities and national laboratories, in terms of industry-funded research. We would see universities and national laboratories – a bit of American language there, I am afraid, for CSIRO, but not to worry – as a major source of new concepts and ideas, extremely important as training grounds. We have heard nothing this morning to speak of about the training role of research in this country, but that is where we see the training occurring – an excellent source of generic R&D support for industry, stuff that is not going to give competitive advantage to one firm over another, and often you get strong tactical support. There are one or two research centres in Australia that are very strong in that area.

Some of the issues that we grapple with all the time? The first one is that intellectual property expectations by universities and national labs are frequently quite unreal. It is startling how many times we end up in months and months of discussion over something that is not going to add much value, and yet the universities can't see it. You have got to have continuous interaction to have effective investment; and you need to make sure you have got agreement on timelines and so on.

Discussion

Francis Rose (Defence Science and Technology Organisation) *I was interested in your 'balanced technology portfolio', in which you had three items, which however have different time scales for delivery, different time scales for completion. I wonder if you could elaborate on how you set the balance, and how that is consistent with a review process which you mentioned earlier was a six-monthly review process.*

Bob Watts What we call revolutionary technologies would typically be more than five years before they go to commercial, five, six, seven, eight years to commercialisation from when somebody has walked in, say, to my office and said, 'Look, I've got this idea.' The evolutionary ones, where you take existing technology and make significant changes, would typically be, say, three to five years out. And the tactical stuff is short term. Is there a clear boundary between them? No. We don't say, 'If it's three years, it's evolutionary,' or whatever. But obviously anything greater than a year or so's time scale is subject to six-month reviews. The tactical stuff is usually funded directly from the people with a problem and you don't have six months to fix it, you've got about three weeks.

Christa Critchley (CRC Sugar Industry Innovation through Biotechnology) *Bob, I want to ask you whether you think the scientific research training model that relates to the apprenticeship-type system that I think has been used or is naturally being used in scientific research training is still appropriate. Do you believe it could be changed or it should be changed, especially with a view to training researchers that are more in tune with the industry and community needs of research outcomes?*

Bob Watts We'll talk about research level people, so PhDs, Masters and so on. Honours students we take in as well. The major requirement for industry is that they are able to work in teams and can work across several areas. That is a major requirement, and we can usually teach it. By switching people every few months, that will happen anyway. By far the greatest need, of course, is to have people thoroughly understanding the care and rigour that is needed to undertake research.

So I don't see a drastic need to change the current quasi-apprentice type mode, except perhaps to have a little more emphasis on teamwork.

Jim Peacock (President, Australian Academy of Science) *Bob, do you think there is much incentive for universities to train researchers for industry, if industry regards the R&D teams as a red-line target?*

Bob Watts I don't think industry sees them as a red-line target if you make sure you sell the R&D efforts in the language that the management understands – which is the great change, I suppose, that we put into our company. I obviously glossed over an awful lot with Real Options Analyses and all the rest of it, but I had to go and learn that stuff and then go and develop methods so I could present the R&D portfolio using that language. And as I said, the budget tripled, having been cut by a factor of six to eight (I've forgotten exactly) in two weeks. We went from 450 or 500 people down to about 100. The guy came and said, 'Right, you're gone in a month.' So it happens that quickly.

If you speak their language you can persuade them that there is value in it. This is why it concerns me when I keep hearing from the university sector a rejection of what started probably as a request and is now getting stronger, really from the politicians, who are only reflecting the constituencies. You're hearing from them, 'We don't see what value you have added.' And it is just wrong to say you can't associate a real dollar value with it. We do it all the time, every time we give out an ARC grant – $100,000 is a regular one. Where did that figure come from? So yes, you can do it.

Your question was whether you can persuade people to go into industry. BHP Billiton is now hiring steadily. I know the Australian context is bad at the moment, but right across the US, if you look at the government labs, they are under constant pressure to show they are producing value. If you look at the industry labs in America, you will find that after those huge cuts in the early '90s they are all building up, but in a very different way from what they used to be. So it is happening.

Steve Gower (University of Wollongong) *I was one of those 100 that were left from those cuts, if you like. I would like to ask from BHP Billiton's perspective, but more importantly from your perspective of driving the research within that: how do you see the effectiveness of the government's CRC program – the opportunities and weaknesses of that program?*

Bob Watts Well, that's a Dorothy Dixer or what! I'm sure a number of people have heard me say in the past that I am not a great fan of the CRC program. BHP Billiton, as you would be well aware, in many cases got badly burnt. Having said that, I must admit that last year I actually signed off on membership of four.

I think some CRCs are extremely successful from the industry perspective, and some are not. And I think it requires very careful involvement from the industry side if you are to get value from them.

Excellence in social sciences in the context of impact on public policy

Dr Valerie Braithwaite
Director, Centre for Tax System Integrity, Research School of Social Sciences, Australian National University

Thank you for the invitation to share some thoughts with you today. I am delighted to be here. My message is a simple one. Judgments of excellence in the social sciences in terms of impact on public policy must be based ultimately on the same criterion used in other areas of scholarly activity – publications in refereed journals and research monographs. There may be other aspects of engagement with the public interest that we would like to encourage, recognise and reward because it testifies to or even adds to this excellence, but fundamentally our established and accepted criterion should remain our bottom line – for appointment and promotion of staff at the individual level as well as for the funding and recognition of budgetary units, or universities for that matter, at the aggregate level. I now want to set out my reasons for thinking this way.

First of all let me explain the varieties of ways in which social scientists contribute to the public interest – and I am using public interest instead of public policy (which is on the program) quite deliberately because some of our most important contributions are through private actors and organisations, and not directly through shaping government policy – and I will say more about that at the end. But for the moment I want to focus on the form that our engagement takes and package the many different forms into three images.

The first is the public intellectual, those academics whose classroom extends beyond campus as they disseminate knowledge and ideas through the media, through documentaries, through websites, newsletters, public meetings, chairing public enquiries, or through communication with more elite and closed groups, as is the case when one becomes a member of a board or committee, or provides private briefings or seminars to outside groups. In an era where the controlling of information is accepted as commonplace – even legal - and spin is on everyone's lips, the role of the public intellectual is more valuable than ever before, and is likely to become increasingly significant.

The second image is the one that I suspect most of you had in mind when you read your program. We might call this the academic guru – someone who has developed expertise in a particular area and is sought out by government to provide an answer to a problem that they have. One might think of the HECS scheme and Bruce Chapman's contribution in turning an

idea about funding universities into a reality – or one might think of the Compliance Model which we developed with the ATO as a basis for their administrative practices and which actually led to the setting up of our tax research centre at ANU.

The third image is one that is relatively new on the scene and one that I expect – indeed hope - is being given a boost through ARC funded partnership schemes. We will call this third image the academic-community partnership. Here community refers to any entity that is outside the university, the emphasis is not really on the term 'community' but rather on the term 'partnership'. While academics are providing knowledge, systematising observations and data, and drawing inferences that will address problems of mutual concern, community partners also are providing knowledge and observations, and testing out the ideas and explanations of the academics for their feasibility and plausibility. Both sides are learning from the other, and in this process, new social science theory is being developed, tested, refined and so on. The partnership is defined by theory building that addresses the issue of concern and problem solving that seeks to implement theoretical ideas in ways that are desirable and practicable. The partnership means that both sides – academics and non-academics – contribute to creating new knowledge, one through theorising, the other through doing, but each using the other to piggy back their next set of objectives.

So much for the form that engagement with the public interest may take. Now let us consider how that engagement occurs, the processes that are involved. I think it is useful to visualise this as a continuum from contributions that are direct extensions of one's expertise (and therefore require relatively little investment or involve relatively little distraction) to those that involve taking time out or making significant detours from one's mainstream activities to contribute to the public interest. Attached to the latter are far greater opportunity costs – most of us would prefer to be writing an article than attending a community meeting on the need for more nursing home beds or trying to understand how penalty regimes for tax non-compliance can be more efficiently implemented; but in both these cases, our engagement is vitally important to the overall mission of universities – to create and to disseminate our knowledge.

Given the range of things that social scientists do in their efforts to serve the public interest – and given my clear commitment to the importance of these endeavours - why am I not prepared to present an argument in favour of assessing the excellence of these activities in and of themselves? One reason is justice, the other is integrity.

The issue of justice comes into my argument in this way. Academic culture rests heavily on an unflagging conviction among its members that it is a meritocracy – that those who are best at creating knowledge and disseminating knowledge rise to senior positions and do so more quickly than others. Judgments about the worth of an individual in this regard are made through peer review. Now no-one has been a stronger critic of peer review processes than I,

but while I believe there are many ways in which they can be and should be improved, it has never dawned on me that we should deny them the pivotal role that they play in regulating our activities. I am certainly strongly in favour of opening peer review processes up so that diverse voices can be heard and outside stakeholders can persuade or speak on behalf of academics who have gone the extra mile in serving public interests, but ultimately the decision about excellence in research that impacts on public policy has to be a decision that has academic merit as its base – that is, public policy contributions must conform to the standard of creating and disseminating knowledge. Peers can interrogate contributions to the public interest as well as anyone else – they can question whether a policy success story is a case of someone being in the right place at the right time, or reflects ingenuity and creativity in moulding current knowledge to suit a particular situation, or whether the ingenuity and creativity came from the bureaucracy itself rather than being academic input. Of course, they won't always get it right – but they have more chance of getting it right than an outside body counting the number of media encounters, or adding up dollars in external earnings.

The second issue I mentioned was integrity. My starting point here is that universities are important as defenders of knowledge acquisition and its dissemination. I am suggesting that to downplay publications as the base for evaluating contributions to public interest is to compromise our collective mission too much. The academic guru was a label I chose with tongue-in-cheek. Once social scientists engage with the outside world and share their ideas, the amount of control they have over how those ideas are used and shaped disappears. The best of ideas can be implemented in ways that are absolute disasters, but after a period of time can be re-introduced with enormous success. In this process, it is critically important that ideas and implementation are documented, if you like separately – so that they can be disentangled further down the track. The researcher has the publication route for expressing and justifying the ideas, and perhaps even setting down guidelines for implementation. The outside partner has the opportunity to do the same from the implementation side. Both sources of data are important, and each can reflect well on contributions that the university has made to the public interest - but my point is that the ideas need to take precedence in any evaluation process, there should be no opportunity for covering them up or sidelining them, or dismissing them because they did not produce the outcome desired by the government of the day. The university needs to be unwavering in its support of the publication of such ideas in the public domain, and this standard is most likely to be maintained if publications remain our single most important indicator of academic merit.

Assessing excellence within our universities to our own satisfaction is one thing, being accountable to outside bodies in terms of our excellence is another. Here I think the answer lies at the level of the budgetary unit rather than at the level of the individual scholar; and

the methodology should be one of storytelling or narratives rather than of counting dollars or media events. Budgetary units that place priority on engaging with public issues should be encouraged to tell their stories. It seems wasteful to demand this of everyone, but targeted evaluation seems both practicable and useful for assessment purposes. The best of these stories can be verified in any number of ways, and if a university is seriously interested in having such contributions assessed as part of their 'excellence package' they can do so through relying on independent panels that can seek verification from stakeholders.

While this process seems relatively straightforward, I have concerns about it happening. In the main this is because it doesn't fit neatly alongside the direction that evaluation is taking. First there is an assumption that detail is the essence of good evaluation and there is a reluctance to go with more holistic and subjective accounts of a centre's activities. A second problem relates to the units used for evaluation. And often budgetary units that have a high policy profile fall between the traditional units - the individual at the most micro level and the topic areas defined in much the same way as the ARC defines them at the macro level. The solution is not the setting up another special category – again that perpetuates the same problem of splitting a coherent body of research in artificial ways, and in so doing, renders the contribution meaningless. Let me give you an example here. In the recent review that was held at ANU, our five academics sat down together to code our papers – we covered seven or more different topic areas – it looked like a dog's breakfast - yet in reality we have a very focused and coherent research agenda. But this was entirely invisible in the research evaluation exercise. What we needed to do was to tell our story of research informing policy processes and policy processes informing research.

Finally I just want to say something about the label – social sciences and public policy– because although it is just a label it can impose blinkers that do not serve future scholars at all well. These days, we are becoming far more aware of the complexity associated with decision making – we talk about different nodes of governance – places within bureaucracies, in the private sector, in the community, local and international (NGOs for example), where key decisions are made that influence the course of events. In other words, if we are worried about carbon dioxide emissions, government may be the least likely point of leverage for academics wishing to contribute to this debate. They may choose to work directly with multinationals, or international environmental groups. Similarly, social scientists who are concerned about trade agreements that deny generic AIDS drugs to Africa and other poor nations may devote their activities to working with NGOs and serve the public interest in this way. Increasingly, social scientists are making their contributions through other decision makers – they are an interest group in themselves forming alliances with some and opposing others, all intent on shaping law and policy.

This again brings me back to my point about storytelling, not just for purposes of demonstrating excellence but for purposes of establishing our credibility in the future. We should be constantly on the look out for our best stories to demonstrate to our colleagues that excellence in research and policy can co-exist and can simultaneously promote better social science theory and better practices of governance. Once the academy can accept that this is so, perhaps our critics outside universities will come to believe that we know what we are doing and can be trusted to both satisfy our intellectual curiosity and support the public interest.

Discussion

Tom Clark (Monash University) *I wonder if it is worth teasing out an aspect of the view of excellence you are portraying. You used the word 'conservative'. It seems to rest very much on track record. As I see one of the key strategic problems for Australian research, we have the problem of the future of Australian research, the future excellence of research if we are subscribing to an excellence doctrine. Academia is the second oldest work force in Australia after farming, I think, and it is only getting older. Just bearing in mind Bob Watts' challenge to talk about future value, I wonder how in a track record-oriented, conservative, excellence regime you are appraising the future excellence – how we can be getting towards that.*

Valerie Braithwaite Is it all right if I answer this in the way in which I look after it in my own research group? I think that when we go to the universities as a whole I feel a little out of my depth, frankly. In my research group I have all young academics; they are all fresh out of PhDs. So it was a big ask to get them to work with the Tax Office, which is what we do, as well as to meet what I set as the performance criteria of three publications a year in refereed journals – that is what I asked of them.

It was hard, as I say; some of them struggled. But what we did – and the Tax Office paid for this, as an example of how these things can actually have unusual, positive side benefits – was that the Tax Office introduced a working paper series and it was much easier for them to get working papers out, because they were published as soon as they wrote them. And that became the base for seminars and then for refereed publications. Every working paper is turned into a refereed publication, ideally.

So I think that our young academics actually benefited. There are some things that are hard for them, working with me in this kind of environment, but they benefited because they immediately could see something that they had written in print, and they had a justification for keeping writing. One of the great dangers for young people who are working in that public

domain is that they do a lot of talking but they don't get the chance to write, which is what they need to learn to do.

I think it has worked quite well.

Bob Graham (Executive Director, Victor Chang Cardiac Research Centre) *I just want to congratulate you on a beautiful talk. It is clear that you adopt the Perutz Model of Management and you are very proud of your people, and I would suggest that that is part of the excellence in your track record. It is not just the publications you produce but it is who you produce. That doesn't need to be a criterion but in your case I think it is and I applaud you for it.*

Valerie Braithwaite Thank you.

Alan Lawson (University of Queensland) *I was just going to take the opportunity, since you had begun to make a number of distinctions in the course of your talk, to raise the possibility that some of the confusion that we have when we are talking about these things arises from the slippage between excellence, value, quality, impact and so on. Those things are not the same, and I think that you were starting to get towards discussing that. I wondered if you wanted to say more about it.*

Valerie Braithwaite Yes, I think that is right. I may be oversimplifying things here, but I don't think it is our excellence that is actually at stake, that government is questioning. They are questioning our credibility, our right to have taxpayers' money, to be excellent in the way in which we all want to be excellent. And I think that's something quite different.

On the issue of establishing credibility – and I am grouping value and all those other things you mentioned under credibility – I think there are many ways of doing it and when you work in different areas you will express that different ways. There are some areas where you can show dollar value, and if you can, why not. It is just that that standard should not be imposed on everyone. And I think it becomes much harder when you look at the public intellectual and, Iain [McCalman], it is hard to put the dollar value on some of the situations that you were describing this morning. But that is where I think the storytelling comes in. The storytelling encapsulates the excellence, but the real point of the storytelling is to improve our credibility. It is to give what we do credibility.

Jim Peacock (President, Australian Academy of Science) *Valerie, we have been talking about the recognition of excellence, and we have kind of been avoiding the withdrawal of support for the not-excellence. I was wondering whether the storytelling has a risk of providing a generic cloak which permits the survival of…*

Valerie Braithwaite Yes, I know the point you are making. It depends at what level that storytelling takes place, I think, Jim. I can imagine a budgetary unit doing a very good job of spinning about their relevance and importance.

I am not sure of the appropriate level. I would be very comfortable telling a story about my research team, for instance, and then providing names for verification. I should also add on this verification issue that our centre is going to be finishing soon, and recently the ATO wanted to do an evaluation to decide whether their money was well spent. And they did do that evaluation. I had nothing to do with it; I decided that it was best to stay out of it.

But it was quite a hard-hitting evaluation. I got a copy of the draft of it, and in it I saw things like, 'She doesn't seem to realise that she shouldn't be biting the hand that feeds her,' and, 'All she cares about is publications.' There were some quotes in there, and of course I stewed on these for days and hours until my husband said, 'Well, isn't that exactly right? Wouldn't you be offended if they were saying anything else?'

Out of that story of conflict and a lot of venting by the Tax Office came [inaudible] where they were absolutely delighted and decided to continue. So I think sometimes we have got to be a little bit brave and put our credibility, if you like, in the hands of those that we feel aren't always sympathetic to our cause. Of course, it is going to backfire at some times in some ways, but I think that is the answer to that question.

Assessing excellence in research in terms of private and public benefits

Dr Geoffrey Vaughan
Chair, Cooperative Research Centres Committee;
Industry Research and Development (IR&D) Board Member

I have got two challenges. One is time, but the other is to discuss issues related to two unrelated organisations, the CRC Committee and the IR&D Board.

I certainly agree with Robin Batterham when he set us off by saying he wants us to 'integrate' everything you hear today so that by the end of the workshops we have got some definite parameters that come out of it. Now, I agree with that, but the difficulty I see is, for those mathematically inclined, the limits of the integration – the area under the curve. The limits could either be, let's say, pure and applied, one end of the integral or the other, public and private, one end of the integral or the other, competitive versus block – it goes on and on – individual versus collaborative. And so that is going to be the biggest difficulty that you face, and indeed it is the biggest difficulty I face in talking about CRCs and the IR&D Board.

Most of my comments will be associated with collaborative research, cooperative research, user-driven research, but even this does have a dependence on individual research at the same time.

Let me just discuss some definitions of what I see being in the public benefits area. I would define this briefly as adoption of research, adoption which will lead to some areas of national benefit – the famous triple bottom line. Generally in Australia it is pretty hard to escape participating under the National Research Priorities. And, even in public benefits, taxation comes into it: taxation at the front end because they are paying for the research, but also at the other end, from outcomes, savings in public expenditure and taxation. A great example of that in public benefit research would be the CRC for Aboriginal Health, where any savings in the Aboriginal health budget have a direct impact on savings in public expenditure down the line.

Employment issues come up, as does export/import replacement. People may not see this as public benefit but it is, because any change that we can have in trade balance allows money for other public benefit activities and that is important as well. Increased skills are very important as a public benefit, and CRCs have done a lot in this, in relation to the research training associated with CRCs. One of the measures, I think, of excellence in research is

reflected in excellence in training – increased skills coming out of CRC, training tomorrow's scientists – and the skills training can't be avoided in looking at excellence in research.

I guess in looking at public benefits one has got to have performance indicators to measure all of these issues and many more. The important thing I think we have got to see, as others have mentioned, is that we don't want this to be all-consuming, we don't want to overdo it and we need to have some performance indicators which do the job without hindering research and research activity. Researchers have still got be left free to roam, there is no question about that in my mind.

In going to another definition, private benefit, under one heading I would put this as commercialisation of innovation. Here we are looking at issues such as new industries, new products, turnover and profit, expansion, efficiency, productivity, competitiveness, technology transfer, income from royalties, licence fees, equity, sale of technology et cetera. And again you have got to look for performance indicators that measure private benefit. To do that in a CRC has some difficulties, but I will try to get to that very shortly.

We are looking at performance indicators right across the board in complex areas. Public indicators are not going to be the same as private indicators, and for CRCs versus the rest, collaborative and cooperative research is not going to necessarily have the same indicators as individual research.

But what can we do with indicators? Well, that is obviously simple and we have all been there before. You can't think of much else other than inputs, outputs and outcomes. The inputs and outputs are very easy, nevertheless they are valuable because some of the value-add you can get directly by looking at input versus output: let's say the number of high-quality PhD students coming from a program. That could be directly found from input versus output.

A harder one – we all know it and we have been through it before with the mapping exercise – is the issue of outcomes. On top of performance indicators, I think the measurement of excellence of research and research activity, particularly in CRCs, comes from reviews. These reviews may well be peer review, they may be in-house review, which I hope is going on in every CRC, with the board looking for evaluation of programs and seeing that their strategies, policies and priorities are leading to expectation of outcome, and there are also independent reviews.

Again, we have got to look here at user uptake in the collaborative research end, as well as the other messages that I mentioned of employment opportunity, training et cetera.

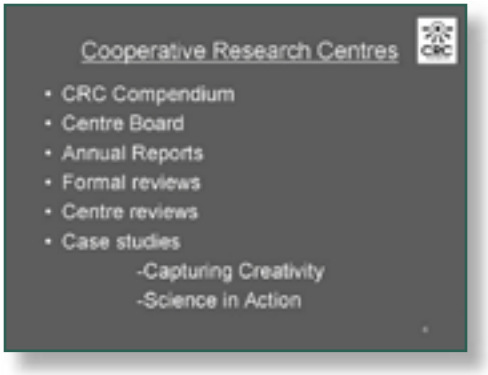
Unfortunately – or fortunately, whichever way you look at it – the logo I am using there [on slide] I have kept on all of my slides, but let me emphasise that there is no direct relationship between the CRC Committee and the IR&D Board, other than having some members associated with both the committee and the board.

When we look at the Cooperative Research Centres we see the challenge. For ease we have got Cooperative Research Centres across six sectors: manufacturing, information, mining, agriculture, environment and medicine. Obviously, you can't have a set of indicators which are going to match all of those discipline areas, all of those sectors. There is a large number of CRCs, 71 at the present time, all listed in our current compendium – if you haven't got one, they have just come out and are available through DEST – so within those 71 you have got a tremendous spectrum of activity, expectation, objectives et cetera, and you have got to be able to have measurement to identify the outcome across those centres.

We have got to be able to measure research excellence at the beginning. In a selection round, when we have many applications, we have got to be able, through the expertise on the CRC Committee – or the expertise, more importantly, on expert panels looking at CRC applications – to look at measuring, identifying and rewarding excellence in research in the outcome of a CRC application.

But it is not the only thing, because we have also got to measure other issues such as participation, user involvement, education, budget – it goes on and on. So although research is important for a CRC, there are certainly other issues that we have got to have.

But we have got a fair bit of experience now. Although there are 71 centres now, if you take every centre that has been established in over eight selection rounds – we are now currently in the ninth selection round – we have in fact started off 144 centres. Some of those are continuing centres, but there have been 144 announcements of a new centre when the selection rounds are added up. They go for seven years; that is 1000 centre-years we will have had by the time the eighth selection round gets through the process. So we are certainly getting the experience. Whether we are getting the mechanism right is yet to be seen.



How do we get the mechanism? I have mentioned the CRC compendium, inputs, outputs – there's a lot of it in there with regard to the funds going in, the people involved, the students involved, the participants involved et cetera. With regard to the inputs, the compendium is excellent for doing that, and certainly at the back it has got not only individual centres but a collection of data and statistics with regard to the input, their research activity et cetera, across the various participants in a centre, whether it be a university, CSIRO or a research institute, government, industry et cetera.

But more importantly, to measure excellence in research of a CRC, we have the annual reports of CRCs. And again, just as we will have had 1000 centre-years, we will have had 1000 annual reports that have come in over the life history of the CRCs, and the ones that are presently up and running. Within these annual reports the centre has to give us a measure of research excellence. Within these reports, under the research programs there are such headings as 'Targets and Milestones Achieved', again a reflection of the research excellence; 'Major Achievements and Outcomes', a measure of research excellence; and, at the back of each annual report, 'Performance Indicators'. Now, this is where the performance indicators for any given centre vary enormously, depending on the centre, depending on the sector, depending on the history, stage and evolution of a CRC. But within there every CRC worth its salt has performance indicators with the objectives, the measurement and a score of achievement against them.

So there are mechanisms in place for CRCs which I think do stand the test for many of them to be excellent measures of excellence in research for any given individual centre.

The formal reviews I mentioned before are important. The formal reviews can be at one stage a review across the system as a whole, and within the CRC history of now about 12 years there have been two outstanding reviews: the Myers review known as Changing Research Culture and the Mercer/Stocker report on greater commercialisation in CRCs. But both of those reports looked at excellence of research in CRCs as well as other issues. They looked at excellence of research in a number of ways, looking at the program as a whole, looking at individual centres through case studies, and they came up with overall a good report on research performance in CRCs.

The other reviews are the reviews of individual centres. Within the program, the CRC Committee carries out reviews of centres, currently at the second and fifth years of their seven-year cycle. We are changing that now to one review at the third year of their cycle; we believe that will be adequate and more useful in the system. So we have got that in CRCs.

The other thing that I think is important is case studies, which give outcomes. The only problem with measuring outcomes through case studies is that you might have to wait a while to see the benefit of the research or to measure the excellence of the research that measures that outcome. Right at the top of the ladder you can see that in the Nobel Prize. Very often, the Nobel Prize is offered years down the track when the excellence of that research has ultimately been evaluated, though it may not have seemed of much use at the time when it was carried out.

Case studies are key and critical. In the CRC program – and this is an attempt to get to outcome measurement of excellence in research – the Cooperative Research Centres Association, along with the Department of Education, Science and Training (or, before that, the Department of Industry) have put out booklets with outcomes of CRC research. It is generally under a heading of 'Recent Highlights of the Cooperative Research Centre Program' but it goes through people, ideas, enterprise, research et cetera, and there have been a number of attempts to tell the story of the CRC program.

Another way of doing it, again through the CRC program – and we have just looked at this in very recent times – is another publication, *Science in Action*. Last year, every week there was a news release of outcomes from CRCs; they have now been put together in one compendium for the year, again telling the story. And from this you can get a measure of excellence in research, as well as excellence in overall performance, of any individual CRC. The big difficulty in telling the story is to know who is going to take it up, who is going to read it. That is one of the problems; we are trying to analyse the benefits of that in looking at telling the story of CRCs.

One outcome was that the rural press certainly took it up. Last year, as outcomes from this document, many items came through the *Weekly Times*, *Land* et cetera, the rural newspapers, and that was one area where telling the story came through very loud and clear.

Let me now just quickly go to the IR&D Board. Like the CRC program, the IR&D program has a whole spectrum of activities. These are the programs which involve research activity run by the IR&D Board: the Start program, the tax concession, the BIF (Biotechnology Innovation Fund), the COMET (Commercialising Emerging Technologies), the IIF (Innovation Investment Fund), the Pre-Seed Fund, the REEF (Renewable Energy Equity Fund) et cetera. They all have a research component to them, and the IR&D Board is interested in measuring research

performance within those components and going about it. But the difficulty is again the difference between the programs that are run by the board, the difference between individual participants in any given program, and the volume of the program.

An example is the R&D Tax Concession. There are close to 5000 companies in Australia getting a benefit at one level or another through tax concession. And that is just an eligibility program; it is not a competitive program at all. It is a straight grant program or eligibility program, and how you measure research excellence across 5000 applicants or participants becomes very difficult. Nevertheless, the IR&D Board is supposed to do something about it.

And it does it through the annual report. As a statutory requirement the IR&D Board is required to publish an annual report, and within the annual report for each and every one of those programs that we have shown there we have got objectives, budget/expenditure, program performance and outcomes. The first three items there – objectives, budget/expenditure and program performance – are basically an evaluation of the IR&D Board's role, the IR&D as such, but nevertheless we have got to report on programs and the individual participants, and that is where we rely very much on outcomes. Throughout the report, again telling the story, there are case studies showing where successful applications, successful ventures, successful outcomes have come from the IR&D program.

So the one thing I guess I am putting forward is that one of the best measures of outcomes which everyone finds is very difficult is, I think, best given as a case study. But again I warn you that that has got to be done in light of time, sometimes, because there is sometimes a lapse of time before the benefit and outcome of an application is found.

On measuring research excellence I should just say that those investment funds – the IIF funds, the Pre-Seed Funds et cetera – have the challenge of measuring excellence. Whereas an application for a CRC or a Start grant or whatever takes a long time, a long evaluation phase et cetera, if you want to go into an investment area and go to a venture capitalist, you have got to sell him your story in 10 minutes. If you haven't convinced him of the worth of investing in your research in 10 minutes, forget it. All the rest of the stuff won't count, because he or she makes up their mind in the first 10 minutes of that encounter. So that is another interest issue associated with measuring research excellence.

Anyway, the IR&D Board is certainly interested in outcome research. Out of the recent agenda papers we can dig out agenda after agenda after agenda on performance measurement as far as the IR&D Board programs are concerned. A current agenda item which took a long part of a meeting was 'Outcome performance measurement'. We looked at issues of what value we were going to get out of that, and such things as executive alert, so that people at the top end of the administration of the program would know if a research grant was going down the drain, if the research wasn't meeting its objectives, if outcomes were unlikely to be met.

We wanted to make certain that the research being carried out was relevant to the program, meeting the IR&D Board objectives, which in the short term are meeting the National Research Priorities and the national benefits as defined by the board. And we wanted to consider how we can best measure outcomes of the very many programs operated by the IR&D Board.

It goes wider than the board because it goes right across the Department of Industry, Tourism and Resources. The department as a whole, I think, is getting a little bit too heavily directed towards performance, because it is starting to become, in my mind, a little bit all-consuming. But the department itself is looking at the external interest in outcome indicators, and there the audit department, the government itself, individual politicians are hounded the department: 'Give us outcomes. Let us know what you are measuring. How do we see whether it is effective?'

There is an evaluation culture being established in DITR, from the Secretary downwards. There are corridor messages going out saying, 'We've got to be able to measure performance. We've got to be able to measure excellence. We've got to be able to measure right across our programs, not just the IR&D program but every program in industry.'

There is an outcome indicator working group. They have got to match their reporting to Backing Australia's Ability requirements of reporting. They have got indicator workshops, matching the department's key performance indicators (KPIs), including the IR&D Board KPI, with what is proposed as Backing Australia's Ability KPIs, so we are not reinventing the wheel – there are no gaps, there is not duplication et cetera. They are setting targets, they are looking at international benchmarking.

So they are going to a hell of a lot of effort to see that they can measure performance across the research activities, in particular, of the DITR programs. I repeat that I am a bit concerned that there is too much energy going into this. There is too much in resources and time being taken up which I think might be able to be spent in better areas.

Looking at it collectively, I reckon the best performance indicator of our research, at any level in Australia but particularly in collaborative research, really comes down to looking at four outcomes.

We are looking at BERD, business expenditure in research and development. That has got to come up in Australia, and it will come up if business will support research – admittedly, applied research, strategic research – in Australia.

We are looking at employment. Through our research we have obviously got to have increased employment across the country and have a good social outcome from research, as could be indicated from employment.

We have got to look at good balance of payments, so that we are not totally consumed with being a Third World country or African republic or whatever they used to call them – that we do have a very high order of economic stability and ability.

And we have got to look at investor confidence. Whether the investors be at the end of the IIF-type funds, the simple investors, or the long-term stock exchange investors, the insurance and superannuation companies et cetera, we have got to have that confidence to maintain a healthy investor confidence across the nation.

I have this theory, that if you look after BERD, employment, balance of payments and investor confidence in the area of collaborative research, cooperative research, industry research, you are doing pretty well.

Discussion

Doug Hilton (Walter and Eliza Hall Institute of Medical Research) *You mentioned that the CRC system is very complicated, in that it is funding a lot of areas, a lot of different sectors, with a lot of different management structures and missions. I am wondering how, with the different reviews that you do and then through different selection rounds, you get consistency of evaluation across your different committees that do the reviewing, and then how you integrate the funding choices that you make across the different sectors when you come to give a list to government in your recommendations.*

Geoffrey Vaughan Yes, it has been a challenge, it has been a difficulty, but we depend very heavily on expert committees. One of the chairmen of our expert committees, Max Brennan, who looked after the physical sciences committee, is here. Max picks people with expertise in the field, they go out and visit the centre, the centre itself has a research survey done with external independence, and the systems come together like that as the second and fifth year review.

However, because of some issues and some challenges and some problems we have reviewed it. We are now going to have just one third-year review, and the review is going to be carried out by the centre. The centre board is going to take the role and responsibility of effective evaluation. It will have to appoint an independent outside assessing team, it will have to appoint its terms of reference to determine how best it is going to measure its performance, including research performance, and it will have to report to the CRC Committee when the review has been taken.

In three years they have had long enough to get going. They have got enough time for us, if we want to, to close them down or to tell them that they are on the wrong track and they have got to do this, that and the other, and do it that way.

Doug Hilton *But for funding new centres, how do you then decide and integrate priorities in physical sciences versus IT versus photonics versus medical, when you come up with your funding list?*

Geoffrey Vaughan That is one of the challenges. It depends on how much money is asked for, it depends on how we see that money as value for money to get the job done, it depends on how much support is coming in from the participants. And you get a feel, in looking at a CRC application, that one has got a better package than the other – one is more competitive than the other, one should be further evaluated through some external evaluation, peer evaluation, expert evaluation, going and visiting them, having an interview. So it comes together that way.

Then we are interested in how they are going to measure their performance to tell us what they are getting out of the program.

Michael Barber (Executive Director, Science Planning, CSIRO Corporate Centre) *Thanks, Geoff, for a good description of the CRC processes. The CRC processes, however, are driven almost exclusively, as you described, by the selection of centres up front. It has been a comment that there has never been a structural consideration of the CRCs towards building the Australian economy in a more holistic or strategic sense. Have you any comment to make to that sort of claim?*

Geoffrey Vaughan Well, the applicants have got to show how they are going to add value, how they are going to participate to the national benefit, how they may be participating towards Australia's National Research Priorities. And I think if applicants have got to do that, they are seeing how they are going to be participating at the best level in the community at large.

Michael Barber *I guess it is partly linked to the previous question. Would it have worried you if we had ended up with all the CRCs in a particular round focused, say, on the agricultural industry?*

Geoffrey Vaughan It would certainly worry me, because I would think that something was wrong with the other five sectors that they couldn't get up excellence applications. History shows that we have never had a ration, we have never had a question of how many we are getting here or there; it has all fallen into place by natural events. We ultimately get a league ladder of applications to refer to the minister – we only give advice to the minister, we don't make the final decisions, by any means, but we give him a league ladder of where they are – and spotted down that league ladder, it so happens, there has been a pretty good spread.

Now, some of them are a bit out of kilter. You would have noticed a lot more in agriculture and environment than, let's say, IT and mining. But I guess that reflects the number of applications you get in those areas as well.

John Byron (Australian Academy of the Humanities) *A couple of weeks ago I was at the CRC Association conference in Adelaide, doing a bit of industrial espionage. I saw a lot of good sessions but the best one, I reckon, was chaired by Michael Barber. It was the postgraduate presentations, and there is obviously some really excellent work going on there.*

But I was wondering if you could tell us a little bit about how excellence in postgraduate and higher-degree research is assessed or viewed in the CRC model, where there is presumably a bit of push and pull between the two different outcomes: the students are sort of trying to straddle both worlds.

Geoffrey Vaughan There is a fair bit of effort put into evaluating the educational component of CRCs, and because the biggest component in education is postgraduate research, particularly PhD research, we hope that it is innovative research and we evaluate it accordingly. We get some measurements because we can look at the number of APRA [Australian Postgraduate Research Awards] grants the students have got, the number of external grants they get, the number of publications they participate in et cetera. We get all those as performance indicators, generally, from the annual reports.

The other thing is the students themselves. We meet with them and they tell us – and they are pretty honest – what they are getting out of a CRC. And very often they say, 'Look, the guys in the lab next door are jealous of us. We have quicker access to equipment, we have supervisors that come and see us, we have generally multi-supervisors from industry or CSIRO participating,' and that all adds to excellence in research training as well.

As you rightly said, the CRC Association plays a role. One insert in these outcomes of CRCs [holds up some publications] is the synopsis of each of those graduate students that win the awards of the CRC Association. So I think people can look at these and you start to see something about the quality of activities of research students in the CRC program.

The CRC students do get a bit extra. They generally have some stuff on commercialisation, research management, even human resource management, intellectual property et cetera, and all that adds, I think, to a better research program – for those people that want it – than some of the other types of PhD programs. Not everyone wants that sort of stuff.

The other thing is the employment outcome of these students. Very often there is a high transition of CRC students to industry, and we are starting to hear this term time and time again, that the students are starting to become 'industry-ready'.

Sue Serjeantson (Australian Academy of Science) *Is that good?*

Geoffrey Vaughan Yes, for those students, not for everyone.

Changing research practices in the digital information and communication environment

Mr Colin Steele, FAHA
Emeritus Fellow, Australian National University

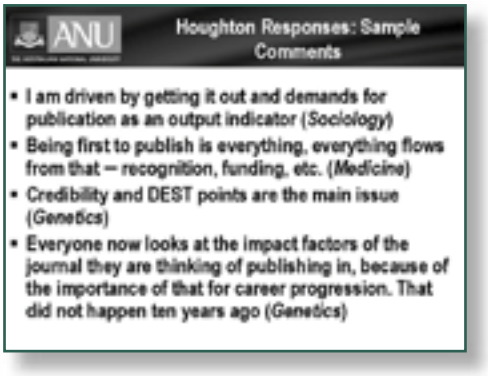
Colin Steele organised a National Scholarly Communication Forum, which was held on 1 June 2004. He hosted Sir Gareth Roberts from the UK, the author of the recent review of the UK Research Assessment Exercise. In this paper Colin summarises the outcomes of that forum.

When I originally was asked to talk I thought it was for 30 minutes, so you are going to get 32 slides in 20 minutes. I would also like to make reference to the NSCF conference, which is extremely difficult to summarise in about 20 minutes for a whole day.

But the presentations by the participants are up on the Australian Academy of the Humanities website (**www.humanities.org.au/NSCF/current.htm**), and I really would refer you to, particularly, Dr Evan Arthur's presentation at the end, which I will come to. I am taking a few slides from each of them, trying to pick up some of the issues in relation to this conference. So do go to the Australian Academy of the Humanities website.

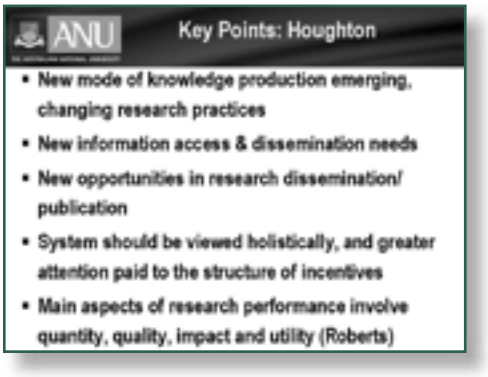
Sir Gareth Roberts is the Chair of the RAE for the UK, and his so-called light-touch research assessment would be extremely relevant to discussions here that will take place with DEST and as part of this forum. We were involved with this as an outcome of a DEST study, *Changing research practice in the digital information and communication environment*. The URL is **www.dest.gov.au/highered/respubs/changing_res_prac/exec_summary.htm** and the full text is on the DEST website. There is a printed copy, and there is also a copy in the ANU Eprints. So it is available for everyone there.

Basically, the key questions from John Houghton were: how do researchers conduct research, what are the major information sources, how do you access and manage information? The whole issue is in scholarly communication: how is it changing in mode to science, interdisciplinary, inter-institutional and global cooperation, and what are the outcomes in scholarly communication and the implications for research infrastructure?



These are just some very quick quotes. John went and interviewed 75 people around Australia; we had focus groups as well. These were leading researchers and some of you in the audience may even have been interviewed by John. All we are trying to indicate there is the changing nature of communication in the sense of the importance of the publications.

And, in defence of DEST, as it understands it the quantitative basis and the publications rating for the DEST points, which are really quite an inhibitor, as Iain McCalman said, was put in by the AVCC and not by DEST. Certainly that quantitative analysis has to go.



Here we have the new modes of knowledge production emerging, new information access and dissemination, the new opportunities for research. What we are arguing is that the system should be viewed holistically: look at the whole process of creation of knowledge, distribution of knowledge and then access to knowledge. Don't let's keep breaking it down into small items of peer review, publication, communication; look at the whole system, what is the copyright, et cetera.

Sir Gareth then came in, in terms of his particular emphasis, and the main aspects of research performance involved quantity, quality, impact and utility.

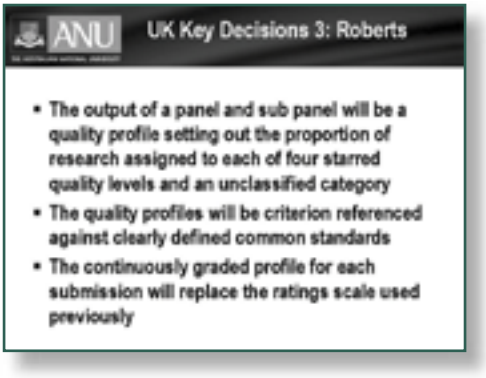
Robin Batterham picked up the three points this morning that Sir Gareth mentioned in the introduction to the National Scholarly Communication Forum. What is the purpose? To allow the funders to assess the quality of research, arising from the investment of public money. I would certainly echo those commentators who said that we do need to get some practical outcomes. And if you have followed, as many of us have, the previous research assessments in the UK, you would know that the way they are moving at the moment does have a lighter touch. You have had the scenarios in New Zealand and Germany coming up as well.

It enables the academic sector to assess its success, informs future strategy and informs funding models. The strategies that have been used historically or prospectively are expert review – and we have seen that expert/peer review, what does that mean? – the metrics, which is going to be extremely important in this presentation of the implications in the publishing industry, self-assessment, mentioned by Iain McCalman, and historical ratings.

The next UK RAE will be an expert/peer review. The census date will be 31 October 2007 with results published in late 2008. Then subsequent exercises will be on a six-year cycle.

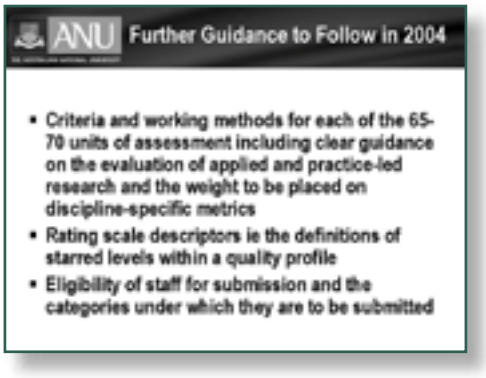
You will see the difference from the previous RAE in the 20 to 25 main panels and some 65 to 70 sub-panels. The questions that were addressed to Sir Gareth included how you can do subjects of departments, which are a composite of individuals, across disciplines – the multidiscipline approach, the multi-collaboration approach and the assessment of standard metrics within that.

Within the two-tier panel structure, arrangements need to be put in place to ensure consistency and the sub-panels need to be encouraged to specify metrics appropriate to their discipline. Some of these wordings are almost like mission statements, but in fact there is quite a lot of detail going to have to go into those, and of course the bureaucracy of the panels that will support them. And there will be papers coming out on metrics and others, later this year.



Other matters were the quality profiles and the four-star assignations, as opposed to five before; the nature of the quality profiles, criterion-referenced. And Sir Gareth worked through some of the issues of how he did that from engineering and the sub-disciplines within that.

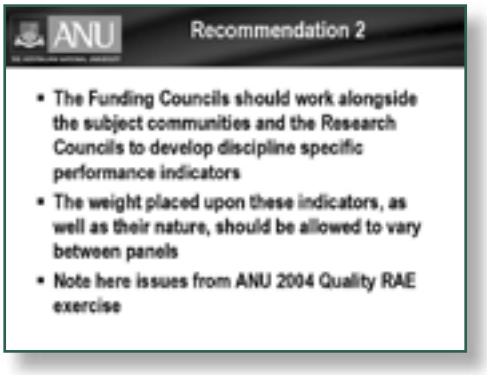
We have transcribed his transparencies into PowerPoints where we can, and they are all on the Australian Academy of the Humanities website (**www.humanities.org.au/NSCF/current.htm**).



Clear guidance is clearly needed on the applied and practice-led research; we have seen the comments about how one does this in terms of industry focus – or, indeed, as he talked about, the 'third leg', people who are working very much with the local and regional communities, such as in the Northern Territory, and how you put the assessment in terms of research evaluation on that.

We come to rating scale descriptors [in second part of slide] and then the eligibility of staff to be included. Sir Gareth alluded to the *Times 'Liar' Education Supplement*, because obviously there had been quite a few references in the *Times Higher Ed* to various universities – Kings College, London et cetera – actually taking people off into full-time teaching so that they would not get evaluated in the research process. And of course some of the New Zealand dialogue came into that: who are the people who will be assessed for their research output?

That [on slide] is their mission statement. You really can't disagree with that, but you need to work through the issues.



This is about funding councils working alongside the subject communities and research councils to develop discipline-specific performance indicators. Again the question is going to be: what are those performance indicators, other than the generic peer review of the panels? Those are going to be very crucial issues. Some of those issues get reflected in the ANU's current quality RAE exercise, in terms for example of where the overseas assessors are dialling-in to the research publications database of the top five. And there are very interesting spin-offs with respect to that.

One of the issues he mentioned which was quite important was the reliance on STM (scientific, technical and medical) metrics, the peer review of peer review. We might actually not read any of the articles because they have been published in this journal and they have already been indexed by ISI. I think there are quite a lot of dangers in that, certainly between disciplines. It is relatively easy, it is argued, in chemistry and astronomy, less so in computing and engineering, and certainly fraught with issues in Asian studies, Pacific, social sciences et cetera.

You may have seen a recent *Economist* article on alleged flaws in the *British Medical Journal* and *Nature*. Two Spanish researchers found that 38 per cent of the *BMJ* and 25 per cent of *Nature* had flaws in the statistical evidence – which only led to 4 per cent of criticism in the entire conclusions, but we are saying again: what is the nature of peer review operating in journals, particularly when that is volunteered, often by hard-pressed academics?

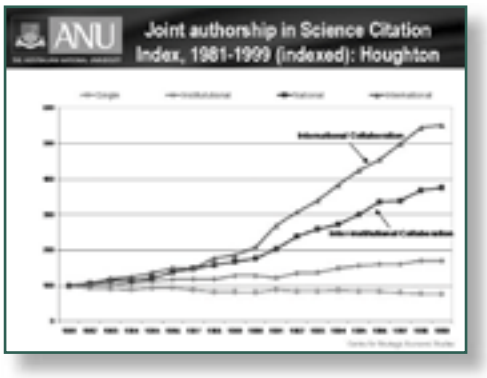
There could be opportunities for new metrics, especially for social science-humanities, relating to the non-ISI citations, and particularly in relation to repositories.

There is also the contrast that Iain McCalman made about the importance of the monograph and the decline of traditional university presses. Professor Blaise Cronin, of Indiana, has just trawled through all the tenure/promotion credentials at the major universities in America – Stanford, Harvard, Cornell et cetera – and in the humanities the key criterion is the publication of a monograph, which is increasingly difficult for a whole variety of reasons. And then that monograph itself may only sell, on average, 200 copies worldwide. So is that a symbol of tenure, or a communication device?

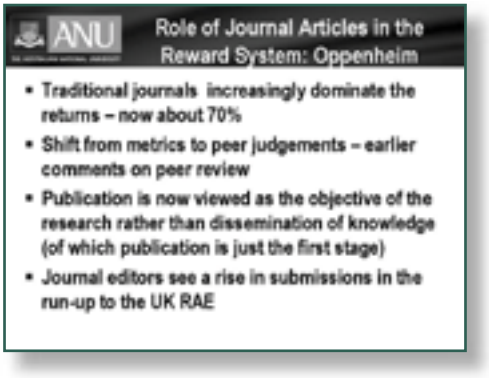
Professor Oppenheim, when he was here, talked about the publications which lead to tenure – what is called in the scholarly communication arena the 'Faustian bargain', where authors give away copyright in return for all the advantages of visibility, citations et cetera – and about citations becoming increasingly important in the assessment of individuals and then, collectively, of departments, and the dominance of ISI, which has really rocketed up in the 1990s since Thomson bought ISI and made it a much more aggressive marketing tool.

ISI is having an enormous influence over the scholarly journal publishing scene. Journal publishers are lobbying hard to have their journals included. For example, only 44 per cent of STM journals, I think, are included in ISI.

There is pressure on ISI to add more open access journals. A new phenomenon called open access journals we will come to later. However, ISI have strict criteria for deciding how a journal gets into their database – again a predominance, as Iain McCalman said, on North America and Europe, and the English language.



This is a graph that John Houghton used in his report, but I just thought I would slip it into the *Science Citation Index* part in terms of where things are going. You are seeing again, as we saw before, the increase in international collaboration as defined by citations within that period, and also inter-institutional collaboration. And when you split that back down again in the ISI-type things, how do you measure the performance evaluation?

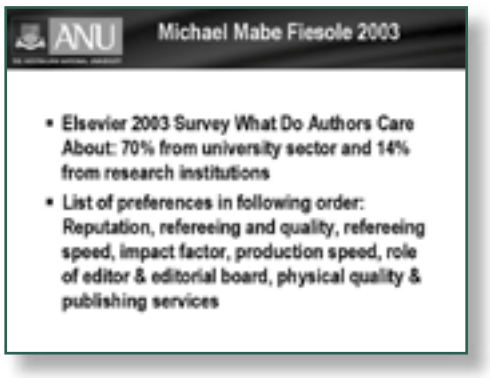


I am just putting those straight up. I know I am going across the thing, but in terms of time it is all I can do.

So again we are getting a reinforcement of the traditional journal arena, and those journals are increasingly electronic – 85 per cent of them are electronic. We are not talking necessarily about print or print output.

We are seeing a shift from the metrics to peer judgment. What he was meaning there was that in fact because an article has already been published and peer reviewed, we don't need to double peer review it. So the metrics are in fact embedded in the peer review.

Publication is now viewed as the objective of research, rather than the dissemination of the knowledge contained within it – which as Houghton showed has often been disseminated well before the publication. The publication is for the accreditation and tenure. And Oppenheim cited that even now, for the 2008, he is getting instructions from his university to start publishing in those journals that get cited. And it may be only two articles, as opposed to five. So people play the game!

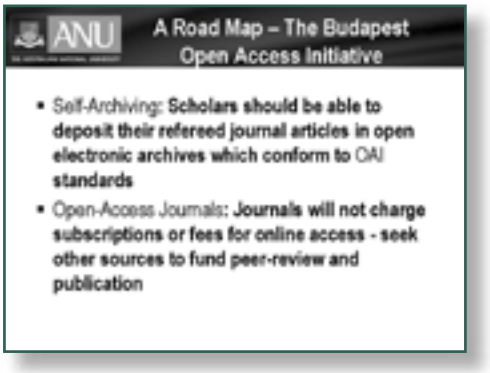


Elsevier, who now are a hundreds of millions of pounds a year profit making multinational, Reed Elsevier, and who publish quite a significant proportion of STM – followed by Springer, Kluwer and others – did a survey in 2003. There were quite considerable differences between the percentages there, in terms of what the academic community wanted. Reputation was the first, refereeing quality, refereeing speed, impact factor, production speed et cetera.

It basically said that authors seek out the top journals, and those authors often don't know by the journal title who the publisher is, whether it is Springer, Kluwer, Reed Elsevier, ANU or whatever. They just know the title of the journal. And then when they become a reader, they change completely from being an author. They rarely want the journal in its totality, but often the article and often free. And none of the financial ramifications of the publishing actually impact on the actual author most of the time – except indirectly, when cancellations have to be made within a university or research organisation. So is this system dysfunctional? And how are we going to cut through those tentacles of dysfunctionality?

We are getting international policy initiatives which you might not be aware of: a declaration in Berlin on open access in October 2003, the Academies of Science in Mexico, the World Summit on the Information Society which had a lot to say about this, an OECD declaration of January 2004, at which Peter McGauran chaired that meeting of the Science Council ministers, largely talking about scientific data but subsequently extrapolated to textual data, and the very, very important UK Science and Technology [Committee] Inquiry into Scientific Publications, which is taking evidence from a huge number of publishers, authors et cetera, research councils, and is going to report in July.

Open access publishing is a bit skeletal at the moment in contrast to the science financial oysters. Elsevier makes a gross profit on its STM publishing of 38 per cent per annum, and we are really talking big business here. When I went to a merchant bank presentation they were saying that investing in Elsevier and some of the bigger multinationals was the best investment around at the moment. But a recent Credit Suisse report said, 'Well, there may be problems with open access,' and Elsevier's share price dipped for quite a bit. It has gone back up again.



The Budapest Open Access Initiative: these are the two criteria and within these there are huge subsets of issues, of course – 'Why should I put it in a repository? I already do in the physics archive,' et cetera. But there are major issues occurring round the world. Self-archiving is one, depositing refereed journal articles or, indeed, guild literature into their electronic archives. And open access journals are ones which are funded by the granting institution or by other means, particularly the American foundations or others. Part of this was stimulated by

the Chair of the Wellcome Trust, in Britain, which has been mentioned: Mark Wolpert funded some major research in Britain, and when he tried to access it last year he couldn't because his library didn't subscribe to the Elsevier journal. He got so annoyed at this that he started supporting the open access movement.

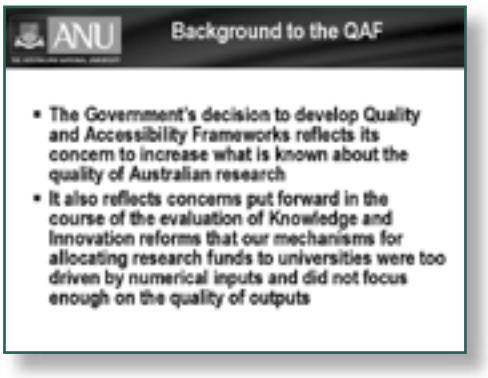
The Joint Information Systems Committee, in Britain, has recently done an open access survey across the academic community. About two-thirds were aware of open access, but only by their institutions, 25 per cent. The primary reason they supported open access was belief in the principles of free access to research information, publication fees to come from research grants, then the institution or the library. It is not actually asking the author to pay, like the author pay charges.

There was much ignorance of the institutional eprint archives, but they said if they were told by their funding authority or their university that open access was a condition of grant, most would comply – again the incentives that Houghton identified.

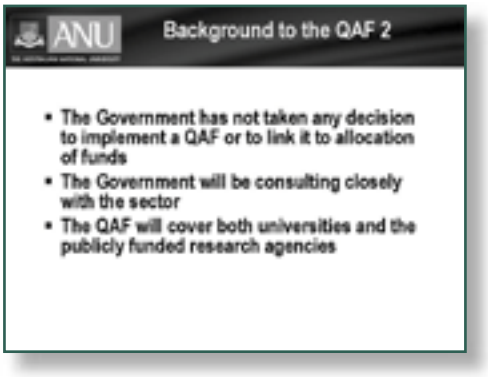
So the NSCF outcomes, before I move on to Evan Arthur's governmental ones, are to encourage the development of a system built on the principles of open access, recognising considerable differences in research patterns even within the sciences, to raise awareness of international trends, particularly in the context of best practice, open access, copyright et cetera, and to look at the incentive systems. Unless the incentive systems change, the practices probably won't, if they remain within the traditional ones.

Dr Evan Arthur's concluding PowerPoints: the next six concluding ones are all his, but his entire presentation is up on the website. This is quite an interesting and very useful presentation, and is probably the most up-to-date presentation of DEST's thinking at the present time.

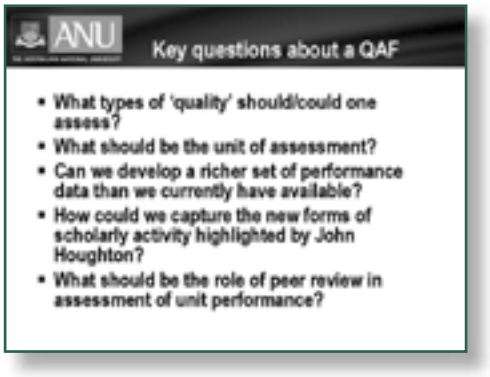
Of central importance are Backing Australia's Ability, national collaborative research, and then most particularly in this area, the Quality and Accessibility Frameworks.



To come back to quality: how do we judge quality, and the quality of Australian research, and the evaluation of knowledge and innovation reforms we have heard talked about? There was also the fact that the mechanisms for research fund allocation were often driven by numerical inputs and not by the quality of the outputs.

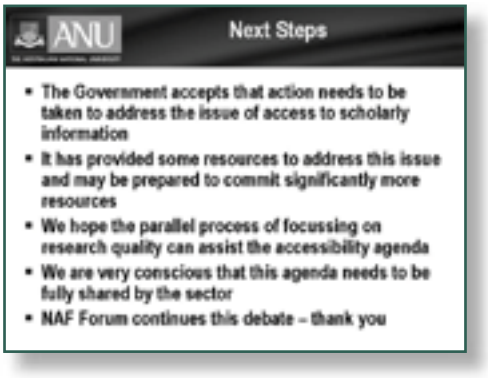


The government has not taken any decision to implement QAF at the moment, nor to link it to allocation of funds. However, the consultations that were alluded to and will be taking place are obviously occurring, and will cover the universities and publicly funded research agencies.



The issues for later on today include: what types of quality could be assessed, what are the units of assessment, how is it going to work, can we develop a richer set of performance data than currently available? John Houghton identified some of the issues that Iain McCalman did, about how you can assess multimedia presentations, the differences in non-book, non-textual, and what should be the role of peer review, as we have indicated before.

We were also talking about making Australian research more accessible, a variation of the Wellcome Trust. There is a major policy interest in improving the accessibility of the research, making it available to all – this is part of this global movement about making material available to the Third World, the developing world – and making research quality more apparent and research results more accessible in parallel, a very crucial point that he brought out.



As to the next steps by the government, there is a need to address the issues of access and scholarly information – picking up some of those world trends that we mentioned before – and government may be prepared (you can see some *Yes, Minister* type phrases) to commit significantly more resources, focusing on research quality and access. There is also a consciousness that this agenda needs to be fully shared by the sector.

This National Academies Forum meeting today plays a very important part in that debate, and what I was trying to do today was to say that there is a very large structural area around here, and things like ISI citations, publications, are a key part – rightly or wrongly – of the research assessment. We need to look at the whole issue of scholarly communication in a holistic fashion.

Discussion

Chris Blackall (Australian National University) *Thanks, Colin, for a very comprehensive survey of this complex area. I did get to see Evan Arthur's slides, and he put up a very interesting slide where for a senior bureaucrat he did something very, very unusual: he tried to sketch a scenario where people would be publishing their work in eprint archives or digital libraries, and the metadata from those libraries could be collected by an agency in lieu of, perhaps, ISI data. What do you think of the possibilities of developing such a system in Australia, given the stranglehold of ISI on this critical data and the fact that it is so incomplete for many, many areas, and key strategic areas for Australia?*

Colin Steele There is a group in New York called the Open Society Institute, funded by George Soros. One of the issues that currently a working party that I am on is looking at is alternative metrics to ISI. That is not to say ISI will go out of business, because *de facto*, Sir Gareth said, they are in fact becoming a metric that is acceptable even though the bibliographical cleansing of their data that is clearly needed, as we saw in the Australian science laureates, needs to be intensively analysed.

So there are a whole series of other metrics that can be used in the institutional repositories, that Professor Stevan Harnad, from Montreal, has actually mentioned and are available. But the problem is that a lot of those repositories are not actually populated in great depth.

Now, ISI as I understand it – and Linda Butler is in the audience, and Vic Elliott is visiting ISI next week, in Philadelphia – are looking at including, as a separate exercise, beta testing of archives, material that is not in the *Web of Science* et cetera, to develop a new model which they can sell at enormous cost. Reed Elsevier have already announced Scopus, which was beta tested at Oxford University, Lausanne and Toronto and which will be a rival to *Web of Science*. So next year, irrespective of these other metrics, there will be two very expensive indices that your university will have to buy, or your research institution will need to access.

So there is going to be a huge increase in the number of metrics. The issue will be, though, whether the government agencies that do the research assessment accept the metrics. A lot of the metrics that are being done by Harnad and others, in Southampton University in particular, are actually very good ones in terms of downloads, digital hits, cross-references, the group call counter et cetera, but one of the problems will be whether they are going to be accepted, and whether they are the de facto metric – which doesn't even need to be analysed – that Sir Gareth Roberts was alluding to. So the UK metrics paper, later this year, is going to be very interesting.

Marie Carroll (Australian National University) *You mentioned the Australian National University's 2004 quality review. Those of us who are struggling with this process – I am one*

of them, and Linda Butler is another – have come to the conclusion that there have got to be better ways of measuring excellence for those who write books. I know that Iain McCalman mentioned this, and you yourself mentioned it just now.

Linda can probably speak for this, but I understand that there are ways in which there is an equivalent of an ISI citation for books, or some institutions are trying to do this now, so that books which, as we heard this morning, might take three or four or five years to have their impact can actually be assessed in a meaningful way, in the same way that the perhaps more peripheral work that comes through journal articles is assessed through ISI now.

Colin Steele Linda Butler is doing a study on book citation in ISI citations. The book one is a really interesting one, because the average sale – and we are talking about research material here, not teaching and learning or textbooks – of a social science-humanities monograph around the world is 350 copies. So again the time of reviewing those books in the review journals is usually pretty slow.

As we have seen from the American study, the book is the symbol of the tenure/promotion, rather than the dissemination of knowledge. What some of the commercial publishers, like Oxford University Press and a number of others, are doing is going to the electronic book with print-on-demand. Each chapter will have metadata, each chapter will have an abstract written by the author – Oxford Scholarship Online does this already – and they will be citing the chapters in the book rather than the book itself in some of their methodologies.

The other ones would be, 'Well if we do say that a book is very important, if there is only one copy of it in Australia and it is in print, how is it going to impact too much on your local community?' So there are major issues to look at here, and one of the things that the Open Society Institute is trying to do is to find the new metrics for social science-humanities.

Linda Butler (Australian National University) *I would just give a brief description of one of the things that we are studying in a current project in our group – which is actually an ARC-funded linkage project, with DEST as the linkage partner. It will be familiar to many of you that if you go in and search for a book in the Web of Science you can actually get citation counts to a book or a book chapter or other non-journal sources.*

The problem in the past has been that ISI does not aggregate those in any meaningful way. All they concentrate on is journals and journal impact factors and journal citations. However, there is a wealth of data in there, and what we are looking at doing is trying to trawl through that data and see if we can come up with some meaningful benchmarks for books and book chapters in the social sciences and humanities.

We hope within the next 12 months or so to have some idea whether it is possible, and whether it can be used as a performance measure or not.

Assessment criteria for excellence in research and research training

Address to breakout groups

by Dr Phil McFadden, FAA (Session chair)
Chief Scientist, Geoscience Australia

Today we have listened to some fantastic talks that have just been stunning in terms of informing us and giving us an exploration about the tremendous breadth and complexity of the issue we are facing here. I think that exploration is going to give us the context within which we are going to be able to move to the next phase of today, the breakout groups, to discuss this question of measuring excellence in research and research training.

I would just like to make a couple of comments before we go to that.

Valerie Braithwaite, in her talk, mentioned integrity quite a lot. Now, in terms of the way people behave, there is an old saying: 'If you don't have integrity, nothing else matters.' I think in the game we are talking about here today, if you don't have excellence, nothing else matters. It is the core, it is the intrinsic need that we have. If you don't have excellence, you are wasting your time. But although it is an intrinsic thing and it is a necessary component, it is not sufficient to give the extrinsic value that we need elsewhere.

The extrinsic factors will be different in the Victor Chang Institute from those in CSIRO or BHP Billiton et cetera. And we need to be able to sort out this inner core: how do we measure excellence?

We all say, and we have said it several times today, 'Well, hang on a moment. You know excellence when you see it.' And there is no question about that. If you listen to Yo Yo Ma you know there is excellence. You know there is not excellence when you watch me trying to run a hundred-metre sprint. It doesn't take you long to figure it out.

We are going to be asked the hard question: how do we actually measure excellence in research in such a way that it will inform funding policy? We want it to inform funding policy, and we want to inform that funding policy in the right way. The issue, then, is: how do we do that?

In many instances it is fairly easy. If you listen to Yo Yo Ma, you say, 'Well, was that performance wooden, or did it soar? Did it transport me?' You have got an immediate measure of excellence. 'Did this guy run the 100 metres faster than anybody else in the world ever has done? Yes? Oh, there you are – excellence.'

In research we don't know what the true answer is, what the real answer is, so it is very hard to judge: 'This guy's response is closer than that guy's response.' So we have to come up with a way of actually measuring excellence, measuring that intrinsic capacity in such a way that it will inform our funding models in the right way.

So what I am going to ask you to do is to go away into these breakout groups and in each of those breakout groups we ask you to examine a couple of questions. In terms of developing a quality assessment framework for research training, what must it achieve? Bearing in mind that we want it to inform funding, what must it achieve? We are going to ask you to add to that some concept of how you would go about achieving that, because that is necessarily going to affect what you think it must achieve. Of course, there is also the issue: what must it avoid? I think it is important to examine that aspect as well.

Similarly, for a quality assessment framework in research: what must it achieve? And to some extent please consider how you would go about achieving that. And what must it avoid?

When we come back here the rapporteurs for each group will have five minutes to give a brief overview of what it is that they have done, and we will then go into a discussion to try and suss out those really critical elements of how you go about developing a measurement framework for excellence, for that intrinsic capacity.

Reports from breakout groups

Professor James Angus, FAA, University of Melbourne (Group 1)

We started with research. We believe that what we must achieve is public confidence, and that there is value in our R&D because what we do in our research sector will apply, as we said, across to defence, security and many other things.

We need some clarity when we are talking about quality versus applied research, the application. We need to develop, within our own sector, our scorecard – whether it be university, CSIRO or whatever – and within that scorecard we have to indicate what are the priorities for that sector.

We spent quite a bit of time talking about the concentration of minds. What do we believe that means? Is it at the department level? When it comes to an institution, would centres make up the quality or the concentration of that institution? This is particularly important if we think about block funding: would there be a university, for example, that may only be excellent in one centre, and therefore that would be its baseline for funding, rather than thinking about whether there should be a baseline across the sector in the universities for block funding.

Then in considering our quality we were interested in discussing where we are going to get the expert advice. And we must not forget the international competitiveness of our research, and therefore of course not be at all shameful or worried about going for an international panel.

That was what we must achieve in relation to research. What we must avoid, we thought, was the 'soft' or the 'safe' research that often is developed from a research fellow who may have one technique and continue to slice that salami for the next 10 years. We want researchers to take risks and develop their own research teams in that way.

We need to avoid the international cringe. If we are publishing in international journals we are doing international research, by that peer group that assessed that research. We have to avoid, in any research assessment exercise, the abuses that occurred after the first phase in the UK. We had a quiet chuckle about whether or not, whatever we bring into Australia, there will be abuses and people will try and avoid (or use) the system. So we need to be preparing version 2 as we start version 1. And we need, of course, to avoid any heavy workload – use the light feather touch – on any research assessment exercise.

As far as research training is concerned, we believe that we must, in our new PhDs, recognise the changed culture of research in Australia and therefore develop in our graduates these

broad attributes. Mainly, of course, today it is around the teamwork, particularly in the sciences but also within the humanities. We don't want the PhD students, because of lack of facilities, to do their PhD at home. They have to come into the university, and one way of warming them up is to get them to teach. So some teaching is terribly important, particularly in the humanities but also in our sciences – to get them into the research labs and the practical teaching labs as we do.

We need, of course, to encourage this capacity to write. It is not only research grants but it is the mini-review that leads on to the important review as it leads into the first chapter of their thesis.

Finally, we really need to look at the better resourcing of our research students. It is not good enough to accept research students in a laboratory at the moment unless you have got other grants, because you are needing those other grants to cross-subsidise. So I believe that is a really important area that Australia needs to look at carefully in resourcing our PhD students.

What we must avoid, I think, is the pressure to complete in four years or less. This is coupled with the 3–3½ years of the APAs, and if we are serious about developing a good PhD graduate we certainly need, in most cases, more than four years to do so. We need to develop opportunities to stop a PhD student, I believe, at the end of the second year and make a very tough decision in many cases, where you might say, 'Let's complete a Masters degree and leave the system.' I think for many students we would be doing them a favour.

Lack of resources again we must avoid. It is not just the wet labs but also spaces for PhD students in the humanities. Many universities are struggling there with the increase in numbers.

And, finally, we must avoid poor supervision. It is all about developing that relationship with a very good supervisor in a good laboratory that builds to that excellent centre which is going to develop so much and be so important in developing the quality of our PhD graduates.

Dr Tom Clark, Monash University (Group 2)

Most of these points are more or less by consensus within the group, but there might have been more disagreement than time allowed us to explore.

We made a couple of preliminary observations, first noting that a pivotal strategy in achieving Finland's outstanding performance on the graph we saw was targeting ISI journals; that has boosted their numbers hugely. We also began with a questioning of the notion of so-called research training. Leon Mann put it much more dispassionately than I would when he said, 'It's much broader than that.'

I am afraid we kind of flowed from one question into another, but I will try and let you know when that happened.

On Question 1, we thought that postgraduate research education should lead to maximal graduate satisfaction. I hope that is pretty clear. The next point in our summary is about ANU but I will come back to that, because actually the ANU has done a major exercise recently.

We have to account for the diversity of higher degrees by research. That is a diversity of modes, a diversity of disciplines, a diversity of ways of interacting with campus and off-campus experience, of industry versus traditional, et cetera.

We must take account of the quality of research that higher-degree research candidates undertake. This goes back to that question of research training. HDR candidates are researchers, and they are trying to prove something about their ability as researchers at the same time as they are learning.

Probably still on Question 1, but starting to fade: we need to accept the use of proxies if we are taking a metrist or measurement approach. I think most of the points here are a bit even-handed about the question of whether we are taking a metrist approach or a peer review dominated approach – the current numbers system or the mooted possible alternative RAE style system – but this is one point specifically about metrism.

Back on research training, so-called: we need to avoid a fixation on completions for their own sake. I can think of an adviser who had carriage of David Kemp's research policies, Andrew Norton, who was a very proud non-completion in political science.

This is definitely moving more generally into research. We need to develop measures and expectations that are calibrated finely by discipline. We need to be clear about the difference between measuring quality, which is a relative judgment, and detecting excellence, which is an absolute judgment. I don't think I was alone in the room, in the sense that if you talk about measuring excellence you really need to deal with a certain paradox of the language there. You can measure quality in ways that you are reduced to detecting or discerning excellence.

We need to support the ethical in research and promote it, and our policy framework should do that.

The ANU's research exercise gives a lot to learn from. It is a big case study, and I think that all of us who are interested in this need to look at what is going on in the ANU right now and into the future, to get a sense of where that takes us. (I am sure the government will be doing the same.)

In 'telling the story' we need to be observant of the needs and aspirations of communities with a stake in our research. Our attention was particularly drawn to research on Indigenous topics and to research with a bearing on disadvantaged communities, with a more general view to the breadth and diversity of the outcomes we report.

We had a few final points. These are getting fairly general, I guess.

We need to avoid simple bean-counting techniques like the bibliometrics – which need not be simple bean-counting but to date have been. We need to promote and assist institutional generational renewal, and we need to pick measures or expectations that highlight this. I think our system should be encouraging departments and institutions to account for how they are preparing for the future of research, not simply how they are standing on their track record of the recent past of their research.

And the last point: for applied research, at least, we should incorporate the end-user feedback better than most of us do. Some of us I think are setting good examples here that we could all learn from.

Professor Neil Furlong, FTSE, Australian Academy of Technological Sciences and Engineering (Group 3)

Many of the points that I am going to talk about have already been covered, so I will be brief.

There are some general points at the start. The group talked quite a lot about the system in general terms: it must be fair, equitable and recognise the diversity across the system.

We also had some discussion about the difference between quality, as a spectrum of activity, and excellence, which tends to suggest the top end of that spectrum. And in understanding that what we are doing here is providing a system of discussion around future funding, we probably need to decide where we want that funding to sit within that quality spectrum.

We spent a lot of time talking about research training and thought that should be quite top-of-mind. That probably has something to do with the balance of people in the group.

To move through to research training: the group thinks that whatever the framework is, it ought to deliver effectively to both student and institutional expectations, and be cognisant of the diversity in the student cohort profile and be flexible in its programs. We have heard about that already.

It must deliver in the expectation of quality supervision – that is a clear issue across the sector. And again to pick up the resources issue, it may well want to look at setting some benchmarks for institutions around minimum resource levels for research students, around balances, for example, in staff time and teaching loads, which often impact on the quality of supervision.

And I think as well the group is talking about the research training framework as recognising the research training programs against the National Priorities and in terms of career progression for the graduates.

We also used those magic two words ‘light touch’; we liked that. And I think in looking at research training programs there were a number of comments around where the responsibility for quality lies, whether it should be sitting in single student-supervisor pairs or whether it is actually an institutional responsibility. I think the latter won out.

To move across now to research: there was some discussion about the diversity of the system and the indicators. I think the group came to a view that the diversity in the sector would be looked at through various weightings across a single set of indicators. And there was some discussion under that heading about an area that has not necessarily been covered in detail today: community impact and engagement with community.

The third point is around a framework that recognises that excellence can flourish wherever it occurs – I think I am responsible for that one – whilst at the same time making the plea for adequate funding for wherever that excellence does occur.

I think the framework also needs to ensure that new starters in the system are not locked out and that they are encouraged to get into the research game. At the same time, however, we need to drive efficiency, across the sector. Maybe that occurs through greater collaboration.

To come to the end of our summary: we talked about avoiding a system that drives homogeneity – I think the diversity in the system is one of our major pluses – and avoids exclusion but again uses the light-touch approach.

Dr John Beaton, Academy of the Social Sciences in Australia (Group 4)

We second everything that has been said.

Taking research training first: we suggest, in measuring excellence of any kind, the use of every measure that you can find, noting that there will be great differences between institutions and even between programs within an institution.

We think it is important to judge supervisors as well as those who are being supervised. We think institutional infrastructure should be judged, particularly as the capacity is highly variable, and note that there are very wide differences in the sector.

We would like to work toward ‘industry ready’ products where possible, and in order to make it possible we encourage industry to start contributing to the university sector so that we can do a better job of it.

We would like to think that feedback from graduates is something that is of particular value when it is five years post graduation rather than six months afterward, when the bitter taste may still be there.

We think it is important *not* to make any assessment procedures onerous, either for faculty or for students. We should try to make reviews a means to promote, rather than punish. And whatever you do, do not force all programs to dance to exactly the same tune.

Now with respect to research: we think the description of excellence must be extraordinarily broad, and we also note that it will evolve. Excellence in universities and publicly funded research agencies may be quite different things, particularly when you take into account things such as commercial-in-confidence research, and these need to be judged in such a way as to not disadvantage individual researchers.

We think that the recognition of differences in measurements within and between universities is extraordinarily important. Again we don't want to see them dancing to the same tune, or trying to. And we think that we should recognise non-traditional means of excellence, not just publication. In other words, people who are developing things which may not reach publication but which have an important impact, either in their university or outside of it, should be rewarded for it.

We think it is important to trust peer review, and I will come to that again in just a moment. Also we think it is important to introduce confidence in the system so that those people who are subject to it are quite happy to be involved in the review process.

We suggest that we do not reduce the complexity in such a way that it makes it terribly convenient for someone sitting in DEST to be able to say, 'The chemists are worth 2.4 and the poets are worth 2.3.'

We say not to manipulate the metrics in order to suit some presumed outcome that someone else may have in mind – a government-driven policy, perhaps – and not to compare full-time and part-time researchers. We know that some people have full-time research appointments and others do teaching and research, so comparing the research output of these individuals is unfair to certainly one of the two groups.

Finally, we don't think it is important to rely on citation indices.

I will, however, take a moment of promotion here and move this a little bit further. In judging research quality assessment, if the goal is for DEST at the top of the following slide to be able to provide funding from top to bottom, and if it is the job of people in the various hierarchical structures to report upwards toward DEST, I would suggest there are a couple of things we want to keep in mind.

First of all, qualitative assessment – that is to say, peer review – is extraordinarily important down here [at bottom of slide] at the review of teaching and research individuals. There is no simple metric, certainly not citation indices, because we know that people quite often have terribly valuable manuscripts that have been accepted for publication, for which there is not yet a citation index, and these things need to be rewarded. There are hundreds of other reasons why you do not want to use those simple metrics.

Similarly, when you are talking about quantitative assessment, where these simple metrics come into play, they are probably more important as we go up the hierarchy here, where measuring quality between universities may be reflected better in things such as total citation indices for the total faculty.

There is one other little aspect about this. We have to move between these levels in the hierarchy, and there is something that we cannot live without here: the concept of trust. If departmental programs, faculties, colleges cannot trust the advice of people who are doing faculty peer reviews, then there is no point having them. If universities can't trust what their deans are telling them about the quality of their programs, then there is no point asking them. So we do have to, we think, encourage government to accept the fact that if they want a qualitative assessment up here [at top level, DEST] by which they can make wise judgments, they have to accept and trust the qualitative assertions of the other levels of the hierarchy.

Dr Ian McMahon, Australian National University (Group 5)

First of all, talking about research training: the group felt that the assessment should look at the benefits, first of all to the student, then to the government and the community, and then to the discipline. So in the sense of the student, what the assessment would be looking at would include the quality of the supervision, the infrastructure, including everything from the facilities – whether it be at an office, the equipment, the desk – to the research expertise of the staff and the supervisor, and then the employability of the degree from the student's point of view and

also the research excellence of the output and the quality of the research that is being done.

For the government and community, that would include the skills that the students have got and their employability, but you would be looking at assessment perhaps from feedback from employers. And then for the discipline: has the research training influenced the discipline? It comes back again to the quality and the excellence of the research.

In terms of some ways of assessing this, you would certainly look at exit interviews. Another group talked about talking to the graduates five years afterwards, and we felt that that was also appropriate, because you get quite a different view of what the student thinks after they have had a bit of experience in the workplace.

In terms of what to avoid, there should be an emphasis on quality rather than quantity, so you avoid purely quantitative measures. You should avoid treating each discipline the same, but also avoid treating each university, each sector, the same. Different universities and different government-funded research organisations, if we are talking about them, have different objectives.

Avoid training for training's sake; avoid blunt, simplistic measures; avoid focusing too much on inputs. (Obviously, part of the assessment will be input-based.) Avoid conflict with institutional management and governments; and avoid too great an emphasis on completions for completion's sake, perhaps looking at the appropriate time instead. At the moment, as was pointed out, students often lose their scholarships after 3½ years and the RTS funding after four years. Rather, you should look at completions in terms of what is actually needed for the degree and set mechanisms and assessments in place that, yes, award quicker completions but only where that is appropriate for the discipline and the work being undertaken.

For research, the emphasis on what we wish to achieve is on excellence. As to how we would measure that, certainly international benchmarking would play a part. That would include the usual measures that we use already, including success in getting grants and continuing to get grants; citations; publications, but with an assessment, perhaps, by an expert group; creativity and innovation – the implications for the discipline, whether it changes the discipline paradigm – industry links and funding.

It was felt that one of the purposes of the assessment would be to make Australia – the Australian community as well as the government – feel good about Australian research and to therefore influence government policy. So you are providing an assessment that enables the government to, in effect, give you more money but to be satisfied that the research it is funding is value for money and meets the needs of government and the community.

In terms of what to avoid, this includes a bias against early career researchers; a bias that would disadvantage emerging fields; something that is overly bureaucratic and too expensive

to maintain; unintended consequences; a culture of risk avoidance so that people are not prepared to do research that is of high risk; and complacency.

Dr Mike Sargent, M.A. Sargent and Associates Pty Ltd (Group 6)

Let's have a look at research training. I think what we want to do is to achieve a framework which recognises the diversity of the paths that lead to a research degree, say a PhD. That is a diversity that varies in accordance with the types of discipline you are working in or the type of research you are doing. It should also recognise the need in many cases for mobility, which is both mobility after you do your research program and mobility in terms of collaborative supervision arrangements.

Most importantly, what we think is that the framework ought to value the quality of the research training, not just the quantity. That will require some consideration of the program itself and the students, supervisors and examiners. We also felt that a big element of research training is the creation of that independent, innovative thinker and so the framework has to recognise that.

In the framework we should avoid reducing innovation within the system so that we all become very vanilla-shaped, and we should try and avoid bureaucracy, because what we are trying to do is create a dynamic training system rather than a paper-warfare system.

In terms of research, we think that it is important to achieve an integrity of a peer review process that is respected and understood. It needs to be international in context and it needs to be respected not only by the funders but also by the research institutions themselves.

The framework also needs to recognise the diversity of the ways in which research is conducted. So when we are talking about the unit of assessment it may in fact vary in accordance with the discipline or transdisciplinary groups that are being assessed. We also think that the process ought to make sure that we don't eliminate potentials for collaboration.

The third point there is that fundamentally, by and large in Australia, particularly with our volatile research groupings, it is seen to be essential that we be able to reassess, if required, somewhere in the middle of the extended period in which the assessment will take currency.

Finally, it is important for the framework to recognise potential rather than simply current status. In that respect we need to avoid too much retrospectivity and keep ourselves thinking about what we want to do rather than what we have done.

The framework will have to be careful to not bias against and destroy niche universities and institutions which might have very narrow but excellent research activities. And, finally, if you were lucky you might be able to find a system that avoided Heisenberg's principle and that would be self-adaptive to avoid the game-playing that tends to fill our systems.

Open forum discussion

Session chair: Dr Phil McFadden

If I could do just a bit of a wrap-up.

One thing I heard throughout each group's report was the cry for diversity, the need to maintain diversity, the need not to have a one size fits all response – the need to move away from simple metric measures as a one size fits all, the need to take into account the different needs of each of the groups. One thing nobody has mentioned is that what we are actually seeking there is moving away from a very simplistic method that is currently being used, which measures quantity, to a much more complex system in order to be able to measure quality.

That has resource implications somewhere along the line, but I guess the thing is that we want this done properly, and if you are going to do it properly you are going to have to accept the consequences of doing it properly.

In the training I heard the need for courage at one's centre to be able to look at their students and see whether in fact they are giving you quality. I remember I once made the mistake of trying to bring a PhD student to a halt at two years, and I still bear the scars, but I think you were dead right that it would be much better for the student had they gone on to doing something that suited them better. We need the courage.

We need to avoid the pressure of an artificial shortening in time for a piece of work that in fact deserves five years to train the student and develop the student into a really self-thinking creature. If it is going to take five years, then sometimes that is worth it and we shouldn't have these artificial impositions that bring it to a halt in some shorter period of time.

We need to encourage and engender mobility between different universities. We need to engender different ways of measuring for the different communities.

So there was a lot of meat in that. In essence, every part of it spoke about the need for being responsive to the requirements of that particular area. One particularly strong statement that I heard from John Beaton, and that I think has got a lot of validity to it, is that need for faith in the peer review. If you don't have faith in the peer review, where are you going to go to get advice about quality? I think one of the things that we need to do is to make certain that we can engender, in the funding bodies, faith in the peer system, faith in the expert advice. That is going to be central to our ability to convince people.

One thing that we were a bit short on, in terms of what was discussed – and it is of course because it is the very difficult thing – is that although we spoke a lot about the generalities of 'We should avoid biasing against early-career researchers,' but the question is how. What method are we going to be able to use to actually judge that this person is excellent and help them, and not bias it against those people? So to some extent we need to think a little bit more deeply, I think, about the 'how'. A lot of what was expressed was, of course, very valuable. But it is not valuable unless we can start articulating some of the measures that we can use to actually start achieving that.

Discussion

Miriam Goodwin (Australian Nuclear Science and Technology Organisation) One thing I also note that we haven't really thought about here and yet is one of the great research challenges in Australia is how this exercise can contribute to increasing the level of industrial R&D being done in this country. How can we increase the level of business expenditure on R&D through giving business a greater confidence in the research quality, perhaps giving business a greater guidance on where the best research is being done that is relevant to their needs, being able to then also allow business people to perhaps validate within their own organisations, 'We are collaborating with Institution X or Group Y because they are shown to be the best in Australia in this field'?

There is a real potential here, and it is broader than reporting to government and it is broader than reporting to the community. It is about being able to really address one of the critical issues that we face.

Bob Watts (Technology Consultant) I think you will find that the industry people know where the top research is, and I think you will also find that the level of research that they support in the country is very closely determined by, if I dare say it, the sort of dollar value considerations that I gave this morning. So a reason why there is not more industrial research done in Australia, I think, is that the management doesn't see the value in it.

Jim Angus (University of Melbourne) I just wonder whether or not this is an opportunity for getting the researchers from the universities either to join advisory boards or actually to be on the boards of some of these companies.

Bob Graham (Victor Chang Cardiac Research Institute) I don't have a lot to add, except that it worries me, it irks me that we don't have much pre-clinical research, especially from pharmaceutical industries, here in Australia. In part that reflects, I think, a xenophobia on the part of the companies that are located over in the United States and other countries. They

don't think we can do it, and they think they can do it better. Again I think it would help if we have these expert panels with people coming from overseas: they can get to see what we have here, and that may filter back. And we may also try and get people from industry to come in on those expert panels, and that may help as well. That's the only suggestion I would have.

Phil McFadden (Session chair) I guess to some extent that is targeting some of the question of 'how'. And I guess there is a question in my mind: do we in this country do enough evaluation, you might say, of programs and institutions, separated from the funding itself? People might want to consider that.

Mark Finnane (Griffith University) I just want to pick up something that was commented on by at least two groups. It is not going to be easy, whatever quality assessment framework we use, to arrive at estimations of quality in research training, but there seemed to be a bit of consensus in a couple of groups that quality was not necessarily related to time of completion.

Let's hypothesise that University A, over a number of years, consistently completes its students in 3½ years and another university doesn't worry about completions, and completes them in 4½ years. If we look at all the other factors in terms of assessing the quality of the outputs in terms of the students' careers and they are all the same, wouldn't we agree, on the balance of things and certainly with fairness to the students, getting them out of the poverty situation that they are in while they are students, that completing in 3½ years is better than in 4½ years?

I am not aware that there is any evidence that taking a very long time to complete improves the quality of research training, and a lot of discussion is going on in Australia about trying to improve the way in which we manage research training in a way that is compatible with shortening the period and getting people on into their research in university and industry tracks, and not getting hung up on the idea that you need a German-style PhD program that takes 10 or 20 years.

Jim Angus If I could just by analogy talk about medical training: you can have a three-year medical degree in Canada, or in Australia now there are four-year graduate programs right through to six years from an undergraduate program, including the one at Melbourne, where as an undergraduate we have one year of research for every medical student. I hope the government doesn't say, 'Well look, you can do it five years or four years at this institution but it is six at Melbourne, therefore we are only going to pay for four at Melbourne, or five at Melbourne.' I think that is a terrible way of looking at this.

I think we have to look at what is the quality of the product and take the individual case in terms of a PhD student to see what it is, why they have taken that extra length of time to get through their PhD. And I think then you will see why there are these differences.

I have got medical students who may take six years because they have moved from full-time to part-time as they have moved through their clinical training. I think we have to allow that flexibility.

Phil McFadden I think you made quite a strong point in your comment there about 'if all other things are equal'. If all other things are equal, and you have given the intrinsic capability, then of course my response is yes, it is better for the student and probably better for the country. As an employer I will quite often look around and have in mind quite specific things that I am seeking, and there are times when I would rather pick off somebody who has spent more time in training and is better quality as a consequence – if it actually does lead to better quality. I take your point: if all other things are equal, then sure. But I think very often other things are not all equal.

Gina Newton (Australian Academy of Science) I would just like to make a comment related to 'all things are not equal'. I think this pressure to reduce completion time may seriously impact on some people who have various circumstances while undertaking their research training. In particular I am thinking of women who may find themselves in the position where they are having children, for example, or having to look after sick relatives or whatever, and particularly people who might go into the workforce and then come back into research training later on in their life when they may have, again, family circumstances that mean that they can't do their training in a brief period of time. I think you have got to take those sorts of things into consideration as well, and not disadvantage those people.

John Byron (Australian Academy of the Humanities) As one of the two former CAPA [Council of Australian Postgraduate Associations] Presidents in the room I think I should respond to that point as well, Mark [Finnane]. As someone who is nine weeks off from submitting a thesis after too long, I couldn't agree with you more and I don't think there is anybody, other than the longsuffering postgrad, that thinks that completing early is a good thing.

I think one of the frustrations, though, is that the three-year cut-off for APAs and the four-year cut-off for RTS are so arbitrary. It is a one size fits all model and clearly is not designed out of any commonsense approach that might fit particular projects, let alone variations between disciplines and institutions and personal circumstances. I think it is the wanton stupidity of a single algorithm to fit every postgraduate/HDR situation that is actually the problem, not so much the absolute numbers. I think that probably most people do agree that finishing earlier is better.

Better doesn't necessarily mean it is higher quality, and I think that one of the problems that the groups might have been trying to strive with is that they could see that there could be some trade?offs in the quality of research undertaken while students are doing their research

degrees, in order to get a rapid completion, and perhaps compromises or less beneficial quality outcomes for the entire research education experience, which might include things in addition to their actual research project, such as training in entrepreneurship, teaching experience, lab supervision and so on.

Phil McFadden Could I just put in one comment, quickly. I hear you loud and clear. I once took over a student who had been 18 years so far on his PhD, and let me tell you, it didn't benefit anybody at all. It is this one size fits all that is the problem.

Mark Finnane I think it is just worth stressing that the RTS framework defines time of completion, not in calendar years but in full-time equivalent years of study. So it does allow, and universities do allow, address to equity issues. For the APA, the scholarship frameworks are much more circumscribed, I think, and John is right about that. But even the APA guidelines allow part-time scholarship funding for equity reasons as well.

Bob Graham There is actually very good data, in at least the biomedical sciences, to show that the longer your training, the better chance you have of getting a good job later on. I would have to dig that out, but I am definite that there is.

Jim Peacock (President, Australian Academy of Science) One thing we didn't discuss much today at all, if at all, was that I think we all would like to think that the framework or system that we put into place will not prostitute the qualities that we all think are very important in measuring the quality of research. And yet they have to be able to be appreciated by people who aren't 'like us', who aren't scientists or science students. That I think is a very important point.

Related to that: so we have this system, we haven't talked at all about then who will make the decisions about how it will influence investment in R&D. I think over the next 18 months that is something that we should have some thoughts about, and try and have some inputs.

Phil McFadden One of the things that I think we do need to bear in mind is that we have in place a funding mechanism that tends to drive things towards quantity rather than quality at the moment. That is a concern that we have heard expressed all the time, and I think one of the things we are going to have to look at in the future, for us coming up with those measures, for us coming up with the 'how', is to help people put in place a funding model that will actually drive it towards the quality that we value and the quality that we think will be of help. We have got to figure out how to do that.

Pascale Allotey (University of Melbourne) I just want to make a comment on the discussions that we have had about the peer review process and how important that is. I think one of the problems – and it has already been raised – is the small pool of that we have of peers to do reviews in particular areas. I, for instance, with grants that I have submitted, have had people

send them back saying they are not happy to review them because they have a conflict of interest because it is in the same area. So what happens is that the grants then get sent to junior researchers, and there is research that suggests that junior researchers tend to be a lot harsher and less constructive in the way that grants are reviewed.

I have, for instance, been able to compare reviews from an NIH grant where the NIH reviewers actually are paid as part of their work to review the grants and you get really constructive, really useful feedback, whereas reviews here, by comparison, are very much about 'How can we not fund this grant?' rather than 'How can we bring it up?'

So I think one of the things that we should be thinking seriously about is some sort of training or some guidelines. We have given [the reviewing of] NHMRC grants to researchers in the department who are straight out of PhDs and have no idea what it is they are doing. There are criteria but there is no real training about what it is we are trying to achieve in the peer review process.

Phil McFadden Yes, I think there is a lot to that. I would point out that I have had a couple of reviews from some young people who could have done with some training in how to review, no question about it. I think we have all experienced that.

The small pool for our peer review was raised several times during the day, and yet I am sure that each of us here ends up doing a lot of NIH and NSF grants, German grants and that sort of stuff, so they are quite happy to use us. I see no impediment at all to our using people around the world, and in fact ARC has expanded its pool to use a lot of Americans, a lot of English, a lot of Germans et cetera. So there is a much broader pool than just the Australian one, but I think you are quite right that this aspect actually does need some training, as to how you deal with that.

Mike Manton (Bureau of Meteorology) Today we have focused a lot on assessing excellence in the work of individuals, but I think that as we develop metrics for looking at research and research training we need to take it into account that you will need different measures for the results of individuals, of teams and of institutions. They are not the same. Similarly, when you come to the reason for the research being done, there has been a lot of emphasis on academic research or research that has intrinsic value; we have also touched on research that has commercial value. But there is a third amount of research done which relates to the generation of public goods, and that needs different metrics as well. So those two things are somewhat linked, and as the metrics are developed we must take those variations into account.

Jim Peacock I asked people in my breakout group to have a show of hands as to whether, if there is an evaluation system and there is a barcode or something that comes out as to the

scores, you would like to wear on your lapel a barcode that was based on the evaluation of your institution, or an evaluation of a group, discipline or cross-discipline that operates mostly within your institution but could be other things, or one that operates at the level of you or your own group.

So I first asked for a show of hands, and I ask it here, for the ones that would like the barcode on the basis of the institution. Now we have the one about its being based really on a group of interacting workers, mostly in the discipline but cross-discipline as well. And then who would rather see it, if they had to choose one, on the basis of their own work and their own group?

The respondents to the last two seem to be splitting the money, but in our group the biggest show of hands was a no-show of hands.

Phil McFadden I think that helps to articulate an issue. One of the things that is going to be quite important in trying to inform a funding model is the very question of whether you should be applying these kinds of criteria at the institutional level, at the group level, at the individual level. And yet you can see here, within ourselves, we haven't resolved that question in our own minds. To ask that is still to ask a complex and unresolved question, and yet this is something that we are going to have to give some guidance on, in order to be able to affect the funding model.

Jim Peacock Maybe the question was wrong. A single barcode may not be appropriate at all.

Mike Manton The answer is that it is all three. A scientist or a researcher wants to be judged as an individual. It is like an individual in a community. You want to be judged as an individual in your family, in your community and in your nation. You don't say that you are just one or the other. And that is what I was saying earlier: it is important that these metrics, as they are developed, take into account that people are associated with different groups. It is not impossible, it is just to recognise that we have to take different indicators for individual components and not try to get one indicator or set of indicators that does everything. It is recognising the differences. And it is important to recognise that those different institutions have to be recognised.

Phil McFadden Do you have any suggestions, Mike, as to how that might be achieved?

Mike Manton Yes. If you sit down and say as an individual what you want to achieve, that is when citations and so on can be grouped together, but with teams and institutions you get much more onto the outcomes: user satisfaction, generation of products, commercial outcomes and so on. And as you move to more institutional measures, that is when you tend not to get citations and so on as being as important as for the individual.

Phil McFadden In several of the groups that I was in, I heard comments, 'Whatever the framework is, it should be non-judgmental.' I was wondering if people in the open forum had

any comment on that, about whether a framework trying to measure excellence should be judgmental or non-judgmental.

Robert O'Connor (Australian Vice-Chancellors' Committee) I would just like to pick up on that briefly and tie it to something you alluded to before. It would be nice if we had a non-judgmental quality and accessibility framework, but you and a lot of people in the room seem to be assuming that there going to be money tied to this. If there is money tied to it, there is a judgment being made. So where do we end up sitting?

Phil McFadden My response is that I personally would like to see instances where we were able to go out and make, I guess you might say, determinations about excellence, not judging but looking at excellence and that sort of stuff, but inevitably the moment you do that there is going to be funding attached to it somewhere along the line. In the long run, to some extent what we are looking at is wanting to inform funding policy, so money is going to be attached, in terms of the way things happen. I think that is inevitable. Bob Watts would certainly tell you that is the way of life.

Frank Howarth (Australian Museum) Touching on that issue: one of the things that was tossed around in our group quite a lot and hasn't really come up in this plenary is whether you can have excellence without relevance. Can you have excellence in its own right, or is it even worthwhile thinking about that at this point, when people are going to make judgments? Can we have a debate that simply talks about excellence, or do we need to roll relevance in at some measure or other?

Phil McFadden Can I bring us back to a comment that Bob Graham made in his talk but that hasn't resurfaced. There is no doubt about it that in the ordinary day-to-day business of science we have to do the kinds of things that Bob Watts is talking about, in terms of being able to show the value. But in the long run the fundamental value and purpose of science – this is what Bob Graham was talking about – is how to make the unimaginable imaginable. Somewhere along the line we have to have measures and have to be able to tell the stories and convince people that in fact that the blue-sky kind of research is worth funding, does in the long run actually have intrinsic value – and extrinsic value as well. To some extent that is the question of being able to accept excellence without immediate relevance.

I would make a point that Pasteur, who had I don't know how many patents – we all know who Pasteur is and he did all sorts of fantastic things, most of it applied science – made the comment that there is no such thing as applied science: there is science, and there are applications of science. He was putting his finger on the fact that no matter the purpose of what you are doing, it is the excellence that matters. If you look at what he did, you see that several times he turned the unimaginable into the imaginable. He separated out the

excellence from the relevance, and what he did was to build a core of phenomenal excellence, most of which then turned out to be relevant.

Frank Howarth Could I make a supplementary comment on that, then. I don't disagree one bit with that, and no matter what, the bottom line is excellence. It has got to be there. But in terms of the real world, we are in Australia in 2004 talking about these issues. We can't shy away from the issues that were raised by Michael Barber or indeed Bob Watts about the interest in mission-orientated or endgame research. It is growing, and the signs are everywhere. So I think we need to spend more time talking about how we add some measures of relevance on to complement that excellence.

Phil McFadden We do have to make certain of the balance. Agencies such as the one I come from are entirely mission-oriented, CSIRO is mission-oriented, and we have to make certain we get the right balance between that kind of research and the blue-sky.

Michael Barber (CSIRO) I will make a comment and perhaps do a disservice to Robin Batterham. I don't think I will betray any confidence if I perhaps reiterate a couple of things Robin said this morning in a bit more detail, and it really does address this question.

Robin has basically put on the table a very interesting proposition: if we have excellent research – in the context he was talking to me about, it was science, but I can take it more widely into excellent research – we will have excellent end-user engagement. That was in a sense part of, I think, the two themes of today: the measurement of excellence but also the focus on the delivery of that to outcomes.

So I actually think the question is a little simpler. I think the question is first: can we, across the agencies and across universities, come to some measure of the funding, which is not driven so much by the competitive funding formulas but is, in fact, in the block grants as Robin said, which does actually say that the science that we do – or the research that we do, taking it more widely – is valuable or is of quality as judged by relatively straightforward metrics?

The only complication in the balance of that framework as it relates to part-time versus full-time, which was picked up, full-time researchers and research-only agencies such as in your own, Phil, or in mine, versus other people in universities, is a danger that we see quantity versus quality. So again interestingly Robin says very much that in the UK it is not about everything, it is a few publications, those publications being seen as being a significant benchmark of that activity.

In that context, then, picking up Frank Howarth's comment, the challenge and the proposition are that if we have, if you like, that assessment of the fundamental contributions of our science background, we will actually see those linkages flowing, whether it is into the

biological sciences, into medical research as Bob Graham was talking about or into the social sciences, with the impact upon social policy.

It is a very interesting proposition he has laid before us. I am not quite sure whether it is a valid proposition, but the first part does seem to me a relatively simple exercise, for us to have agreed in some broad sense that our output is able to be assessed in some metrics which are relevant for our own discipline.

I say this because the other point that he made this morning was that in the UK the metrics, the value you apply on that end of the spectrum to what is regarded by the social sciences as high-quality social science research and what is similarly regarded by the biomedical sciences need not be the same thing. We ought to relatively, between those frameworks, be able to differentiate. Then I think the relevance will be an interesting question to actually see, because that I think is more about the corpus, if you like, the whole discipline. What is the impact of the social sciences as a community on social science policy? What is the impact of the earth sciences on exploration et cetera? And we can tease out those dimensions.

Phil McFadden Something I would just stress – I think it was something you actually said in your talk Mike – is that within the mission-oriented science you are given a goal, you know that you should be focusing on that. But curious people, people who are excellent, still find the opportunity to do absolutely stunning science within that, and add an enormous value as a consequence. I think that is critical.

Don Price (CSIRO) One of the questions that has confused me a little through this whole discussion is about time scales and timetables for measuring excellence, and whether we are focusing on measuring the excellence of science that has been done or on measuring the excellence of proposals for science that we intend to do. In many cases there is a relationship between somebody's track record and the science that they are going to do, but in many cases it may be different. They may be quite different questions and the answers may be different according to when you measure them. If you try and measure the excellence of a piece of science immediately it has been finished, then the metrics and the recognition of it are going to be quite different – and indeed the apparent relevance of it might be quite different – from if you measure it two years later or something. If you are trying to make a judgment before a piece of science has been done, indeed, at the proposal stage, then the way of your assessment of it might be again a completely different question.

Phil McFadden I think you are quite right. When Geoff Vaughan was talking, he touched on that question of pre-judging because you are looking at a CRC that you are about to give money to, and making a pre-judgment as to whether there is going to be excellence, as compared with the issue of post-judging some science. You came up with quite different criteria there, Geoff. So Don I think you are dead right; there is quite a significant issue there.

Robert Cribb (Australian National University) I think when we talk about relevance as a criterion for judging research excellence we have to bear it in mind that relevance is not to be judged simply in terms of economic or technological advantage. Research, for instance, which informs public awareness of Middle Eastern politics or of Indonesian politics or of the PNG economy is vitally important. Its value is not quantifiable but it...[inaudible].

Phil McFadden Bob Watts touched on that with regard to Ok Tedi. It is the kind of thing that, although you don't apply a value to it up front, you find has enormous negative value to you if you don't get it right, and so it is something that it really very important to deal with.

Jim Peacock I can't see why we would want to bring this framework in if it isn't judgmental. What do we want to do? Do we want to pin something up on a wall? Why do we want peer review? It's judgmental. Perhaps what people are saying is that we don't want the department people to make the decision about someone being fired from a university. Maybe that's it. But what we want to bring this in for is to optimise the investment of government moneys in R&D and research training in Australia. How the judgments are made, and where they are made, is of course very important. But we do want it to be judgmental. I can't see that it could be anything else. Furthermore, we want it not only to optimise but to be able to be used as an argument for increased support of R&D, to show why it is so important for Australia.

Bob Watts Let me comment, then, on the political aspects et cetera raised at the back there. In fact, BHP Billiton does put a value on political and country risk. And it is expressed in dollar values. If we want to invest in, let's say, Indonesia at the moment, the company requires a much higher effective rate of interest on the money that it is going to put in than it does if we are going into a nice safe country. So there is a value associated with the sort of information that we get from the political scientists and so on. It is absolutely built in to all their financial modelling. And there is a large group of people doing it.

A comment on quality as well: one of the rapporteurs said that excellence is an absolute measure. I disagree strongly with that. A simple example would show you why. How do you assess excellence? Let's go back to our music, mentioned by Bob Graham, who catching me in the corridor said something about Mozart. I don't quite know what my teenage granddaughter would do, but it wouldn't be repeatable, if we tried to say that Mozart was better than some pop guy who, in her opinion, has excellence in music. And our tastes and our assessments change over time. So we do have measurements of excellence that change, are not absolute, are relative and, frankly, are assessable.

Tom Clark (Monash University) I want just to elaborate what that point was. As I understood it – I didn't raise the point – it was that quality is a sliding scale from 0 through a range of fractions up to 1 as being perfect quality, whereas excellence is a value of 1 or 0, with no

intermediate scale. You are excellent or you are not, whereas you have a varying amount of quality. That was the context in which that distinction was being drawn, and I am sure you can draw it in any number of other ways.

Phil McFadden I think we have finally quantified the qualitative! We have got some fascinating discussion going on, but we have reached 5 o'clock and I think we are going to have to draw it to a close because there are people who have to catch flights and that sort of thing.

I would like to emphasise that this does not bring this process to a halt. This is part of a journey. Today we have explored – and I think we have had a wonderful exploration – the complexities, all the different facets of the different things that we are going to have to think about. We are going to have to be involved now in drawing this together, and then in thinking about how we achieve some of the things that have been articulated here.

There have been some very strong statements made about what it is we would like to achieve, and certainly some very strong statements about what we would like to avoid, and I think the next step is for us to start thinking about how we actually go about achieving that.