

**ECONOMIC DEVELOPMENT AND INFRASTRUCTURE COMMITTEE**  
**Inquiry into Improving Access to Victorian Public Sector Information and Data**

Melbourne — 27 October 2008

Members

Mr B. Atkinson  
Ms C. Campbell  
Mr P. Crisp  
Mr D. Davis

Mr B. Tee  
Ms M. Thomson  
Mr E. Thornley

Chair: Ms C. Campbell  
Deputy Chair: Mr D. Davis

Staff

Executive Officer: Dr V. Koops  
Research Officer: Ms Y. Simmonds

Witnesses

Professor J. Rosenberg, Deputy Vice-Chancellor, Academic, and  
Ms A. Horn, University Librarian, Deakin University.

**The CHAIR** — Welcome to the committee. This is an all-party parliamentary committee, and we are hearing evidence today on the Inquiry into Improving Access to the Victorian Public Sector Information and Data. Your evidence today is protected by parliamentary privilege, but comments you make outside the hearing obviously are not. Could I ask each of you, please, to state your name, and if you are appearing in a business capacity, your position within the organisation, the name of the organisation and the business address.

**Prof. ROSENBERG** — John Rosenberg. I am Deputy Vice-Chancellor, Academic, at Deakin University.

**Ms HORN** — Anne Horn, University Librarian at Deakin University.

**The CHAIR** — This evidence will be taken by Hansard, put on the internet and will become public evidence.

**Prof. ROSENBERG** — Thank you very much for the opportunity to meet with you. Deakin University was very keen to make a submission to this inquiry because we think that access to information and access to data is very important for universities. I have a strong belief, and I think most people in universities have a strong belief, that universities are fundamentally about teaching and research. Both teaching and research involve access to data and access to information, and this was the reason why we felt we needed to make a submission. But there is a particular message too that I would like to convey, and then I will ask Anne to speak.

Our mission statement at Deakin is a little bit of an unusual mission statement. It begins:

Deakin University aims to be a catalyst for positive change in the individuals and the communities it serves.

Then it goes on to talk about some other matters. We take this very seriously. We do believe that we have a strong obligation and a role to work with the communities we serve, and in fact our latest strategic plan, approved in April, is entitled *Delivering Effective Partnerships* and has a very strong emphasis on working in partnership with government, employers, industry and the community. We believe we can best achieve our distinct mission through the development of strong partnerships.

I assume that all the members of the committee will be aware of the current Bradley review, and in the briefing paper for the Bradley review there have been a number of mentions, and in fact more generally in the higher education sector there has been a lot of discussion, of three strands of activity for universities: teaching, research and what is sometimes called knowledge transfer and sometimes called community engagement. I suppose we have a very different view.

When I say 'we', Deakin University has a different view about this. We do not see three strands; we see two strands: teaching and research. We see this strong obligation for community engagement or knowledge transfer, but we see that happening through what you might call engaged teaching and engaged research; that we undertake research activities that are integrated and engaged with the community. I think trying to separate those three acts activities actually loses the point.

This is quite a different view, and when you start to think about particularly engaged gauge research, then immediately access to public sector information becomes critical in order to develop that engagement and be able to undertake particularly community-based research. That is one of the themes in our submission and I think a very important theme.

With that brief introduction I am going to ask Anne to speak as well. Anne really has a much greater knowledge than I do about information.

**Ms HORN** — I am just going to go through some of the key points in our submission and highlight those. You will have picked up that our submission strongly supports an open access

model for Victorian Government information. We have suggested that it is difficult to contest the social and economic benefits of an open access model and that there is growing community expectation for information to be freely and excessively available on the internet.

For Deakin University we support strategies that, as John said, potentially enrich community and civic engagement and particularly extend opportunities to undertake what we refer to as engaged research. Improvements in access, to social, medical and scientific research are of particular benefit to us.

The current model of access — and I think the discussion paper has been very honest about this — requires anyone who wants government information to know it exists in the first place, and then you have to know who to go to and ask permission to get hold of that information, which is not always transparent. There are various stories; our researchers tell us about the difficulties they have had in seeking permission at times.

An open access model has been adopted elsewhere, and the discussion paper mentions the Queensland State Government. It is being consistently promoted through the Commonwealth Government's accessibility framework, which is a framework that as universities we are very keenly watching, are part of and which in our own view could only benefit Victoria.

The University recognises that the model adopted will require exceptions that ensure compliance with commonwealth and state legislation and protect the commercial interests of government agencies and their partners. Therefore we have suggested a hybrid model for open content licensing that combines a Creative Commons approach with tailored licences as required to meet some of those obligations. The key benefits of a hybrid model not only lead to discovery, but the first is it makes discovery easy over common search engines and therefore available for reuse, but also the efficiencies in terms of that licence model for those drafting and managing licences.

I note here that the *venturousaustralia* report, which came out subsequent to this discussion paper and is the response to the Review of the National Innovation System, recommends under recommendation 7.8 that:

Australian governments should adopt international standards of open publishing as far as possible.

And further:

Material released for public information by Australian governments should be released under a Creative Commons licence.

Our submission is consistent with that recommendation.

There is no doubt that our researchers look to the US with envy, where government information is not copyright. Certainly the national institutes of health have brought many benefits — no-one can contest that — to global communities, as well as social and economic benefits.

Our researchers are keenly interested in the outcome of this inquiry. They are anticipating improved access to contemporary and historical datasets and research outputs that will promote greater knowledge production and application, bring benefits to teaching and learning, and enrich community engagement.

**The CHAIR** — Thank you. I want to begin by firstly congratulating Deakin University on its approach to teaching research with community engagement as part of that. It has really distinguished Deakin in a way that for me, as a person very interested in social policy, has set you out as absolutely exemplary in your field.

We have had quite a bit of discussion on economic issues, and I want to indulge myself a little in social policy, because I know you do that so well. Let me give you an example. Say we have got a piece of legislation coming into the Parliament or you may have researcher, a PhD student trying

to identify what is good practice in, say, foster care, in permanent care, in birth certificates for adoption or birth certificates for assisted reproductive technology and the children of, as they call themselves, the donor generation.

There compactuses full of this information in DHS. It goes up on the internet for a short period of time and then it is taken off. If you were a PhD student or masters student or even an honours student at Deakin and you were trying to find that kind of information so that you would be able to benefit in your own research, how do you go about doing it?

I can give you that example because we had a piece of legislation last sitting week where I attempted to get this information off the DHS website, and it just is not there anymore, but I know from past experience that there are compactuses full of it. So can you run through the kinds of things that you would require of government to enable your students to benefit and, I would then argue of course, the Parliament and the wider community to benefit?

**Prof. ROSENBERG** — Do you want me to begin?

**Ms HORN** — I might just reflect on some of the conversations I have had with researchers. If you are a PhD student you are relying — at the moment as they refer to it — on contacts for knowledge. There may be some information up on the website; it may not stay and often it does not stay. For a PhD student you are working with your academic staff member and relying up their contacts with the department. And there are good contacts with the department. That is really part of what they do — try to foster those relationships so they can get the information. That is from my discussions with some of our researchers.

**The CHAIR** — But that goes to knowing it is there.

**Ms HORN** — That is right; that is the whole model, as far as I can see.

**Prof. ROSENBERG** — It is extremely difficult. Particularly PhD students spend inordinate numbers of hours trying to find this sort of information. We have people within Deakin — particularly our dean of the health faculty, who used to be the head of DHS, so that helps; he understands a little about where some of this data might be hidden. But it is still often in a form in which it is very difficult to analyse.

**The CHAIR** — We have got to come up with recommendations. If you were writing a report, what would be your recommendation or recommendations in this regard?

**Ms HORN** — There are a number — the first thing is ‘findable’. That means the data needs to be described in a way that can be findable.

**The CHAIR** — Via where?

**Ms THOMSON** — You would have to use a search engine?

**Prof. ROSENBERG** — Yes.

**Ms HORN** — If you want it to be widely out there in the community you would have to find a reliable search engine. It can be findable via the website, but you still require a knowledge. It is really out there in the public domain, requiring that knowledge of where to go and how to manoeuvre through how a government is structured in order to find that information.

**Mr THORNLEY** — The metadata investment is the key, is it not?

**Ms HORN** — Yes it is.

**Prof. ROSENBERG** — Exactly. And there is a real cost there, I understand that.

**Ms HORN** — And there is data that you may not want to be findable that you could still describe. I mean, you might not want to make it accessible but you could still describe it and have it known that it is available.

**Ms THOMSON** — Or not available.

**Ms HORN** — I think the first thing of the commitment to having the information known is that there is an overarching principle that we want the information to be discoverable. Whether it is able to be accessed and reused is another matter. But you do know that it is there.

**Prof. ROSENBERG** — There is an investment required, but that investment has huge value for the government as well. As you say, you were looking for that data and it was not in a form that you could analyse it. I think there is a wonderful opportunity for researchers to provide that sort of analysis that could be used then to develop new policy. But at the moment it is extremely difficult to do so.

**The CHAIR** — Thank you.

**Mr CRISP** — At the table on page 8 you have looked at some of the areas of interest to the University. As I read those I see we do not actually get to some of those nitty-gritty difficulties. I know we have talked initially of it being about just finding it, but is there anything more you want to add to those examples about the difficulties?

**Ms HORN** — I think I said over the page, on page 10, before I get back to that, that I think if there is a principle of open content, it is then about making it clear what the permissions are. That needs to be transparent as well. What is the access? What are the questions? What are the rights to access that information? I think then you will need to look at what data — I am fully aware that there is data that is readily available to be made public and data that is not.

There is a cost to making data available. So there needs to be some sort of analysis of the data that is there and what its readiness is — whether it is a data audit or what it is. This is if I was looking at our own house. I think that is not dissimilar to what the Australian Bureau of Statistics has done on its own — and I know it has been ahead of this for a long time, but it has recognised data that it can make publicly available at no cost and data that is available at a cost that is negotiable with the universities.

The examples here were examples from particular researchers. I have to say they are probably researchers, those who have relationships and those who may have some level of frustration. I think example 4 is an example of the level of frustration in being able to get access to raw data. A lot of the feedback from researchers is about access to raw data and the emphasis that they are trusted. I think it is that they can be trusted to work with that data, to share knowledge, develop knowledge and enrich the research in that area if they have that access. I am looking at some of the other problems around this. I know with the health data it is the de-identified data and that is well understood and that is why I talk of trust. These are folk who understand those issues.

**Prof. ROSENBERG** — Perhaps the other comment to make there too is that I think one of the opportunities is that universities have the ability to take some of this raw data and develop some of the metadata and the structure and then make that available to other researchers and to government. So I think a great opportunity exists there but at the moment finding the raw data and then getting it in a form that is usable is extremely difficult in some cases.

**Ms HORN** — And I have to say researchers do see themselves as working in partnership with those in government in doing this.

**The CHAIR** — Take that disability research example 1 on page 8, do I take it that that has been difficult to access or is it just the kind of information that they find valuable?

**Ms HORN** — That is just the kind of information they find valuable and that is a current example where it is working well.

**The CHAIR** — I was going to say because that is all publicly identifiable. Thank you.

**Mr THORNLEY** — I suppose I just wanted to follow this general discussion about metadata for a little while since we are on the topic. We will be making a number of recommendations out of this and whilst a lot of our conversation has been about the licensing regime and Creative Commons and so forth, and then we have had a lot of discussion about the economics of particular issues around pricing, some of which you heard not long ago, it does seem to me that the other issue is what is the right balance of effort required from government with a finite set of resources to get the maximum impact from whatever it is we choose to do and recommend.

It probably reflects the 10 years I spent cataloguing the internet but I have a suspicion that the creation of the metadata layer is probably one of the more useful things that we could be doing but it is very expensive.

I am wondering if you can just put yourselves in our shoes for a second and give us your thoughts about how you would balance that priority versus others, and what types of metadata construction you think would create the most value most rapidly? At a simple level you just get stuff that is integrated, so Google's spider can find it and then get out of the way right through to a much more active role.

**Prof. ROSENBERG** — I might make an initial comment. Particularly in some of the areas that we have identified long — health, population, environmental change — there is an enormous amount of information and data stored within government that cannot be analysed at present. Therefore there has to be huge value to government as it develops new policy in having better access to that data. I think the opportunity is for partnership between government and research organisations to help develop that metadata. It is expensive to develop but the return is there a thing for both organisations. I think that is the way forward.

The form of that metadata is not my area of expertise but I do know that the raw data is really not in a form that can be analysed in any sensible way, and that a lot of the work to develop the metadata will be very difficult and very time-consuming. But we need to start now because many of these sorts of projects need to be longitudinal projects over a number of years, and the longer we delay in getting this data into a form that is properly analysable I think the harder it is going to be.

**Ms HORN** — I think it is the knowledge base of bringing that knowledge together. If you are to describe the data you have to have a good understanding of what you are using and what its value is, and as you work with that data you understand more what its value is, too.

**Mr THORNLEY** — Yes.

**Ms HORN** — In terms of the metadata, how much you want to spend on that is as long as a piece of string. Again, that is why I started with description. At least if you know what is there and even keep it at a high-level descriptive level, it is an improvement over where we are now.

**The CHAIR** — Thank you.

**Ms THOMSON** — It is also a big task. I noticed the collaborative arrangements that you already have in place for the research that you are doing, and given the enormity of the task that you are suggesting government should undertake and the way in which you might want to receive material and data compared with someone else who might want to receive material and data, it could make for a very interesting hundred years of preparing the metadata for any purpose for which it can be used.

**Ms HORN** — You are right.

**Ms THOMSON** — I guess I am interested in areas of priority. If we are going to get serious about the areas for which being able to find and source and at least, even if it is not in a form that it might be immediately what is required or useful, how it can then be developed in that way and how we could perhaps rationalise the costs in some way to ensure that the government does not bear the burden of what sounds like incredible costs to me in getting all this material into a usable set of data which may or may not be used, and how often it may be used and what the ongoing benefits may be of using it.

I would be interested a bit in and how you might prioritise the areas of data that would have community benefit at the end of the day, not necessarily commercial benefit but maybe that is well, and talk a little bit around that licensing aspect that you mentioned in the hybrid model around the potential return to government for those commercialised opportunities or whether the government should not expect that. I guess what I am trying to do is to tease out what is an incredible task you would set for government and try to put some parameters around it.

**Ms HORN** — As I said, I think it is difficult and I fully understand it is a very large endeavour. We could easily say health because that is a core area. We have strong interest in the social areas — and crime has come through there — but I am just thinking that that can be quite different segments, quite specialised areas so what do you prioritise and is of the most value to society across all the different research that is being done?

I am just reflecting upon the ties that are now put on research within universities that when you undertake research that you undertake to make that data and that research output as much as possible publicly available, and again in the same environment where you undertake a partnership or work with government, that you undertake also that the data and the research output.

**Ms THOMSON** — You are saying this would be an opportunistic exercise, so you would do it opportunity by opportunity?

**Ms HORN** — Which is really the way the Australian Research Council is approaching it.

**Ms THOMSON** — Yes.

**The CHAIR** — The ARC obliges?

**Ms HORN** — Strongly recommends.

**The CHAIR** — ‘Strongly recommends’. Deakin does this anyway, does it not?

**Prof. ROSENBERG** — For most of our research, yes.

**Ms HORN** — Yes, except that which can be commercialised.

**Prof. ROSENBERG** — Which is not a large amount.

**Ms THOMSON** — No, it is not.

**Prof. ROSENBERG** — Universities always have unrealistic views.

**Ms THOMSON** — It is not as much as Monash had.

**Prof. ROSENBERG** — No.

**The CHAIR** — Can I take you back to page 8, example 2 — health population data linkage — where you have given that as an example of government information that would be of value to research. Could you expand on that a little more? I thought that was a really interesting

example. Within our report we are trying to come up with examples that work well, value add, which highlights the point you have just made with Marsha Thomson.

**Ms HORN** — It might be best if we send you more information on that because there is a lot of information behind this sent to me by different researchers. I would not want to misrepresent what they have said.

**Prof. ROSENBERG** — We could certainly provide some detail around the sorts of research and the sorts of data access that would be required to undertake that research. But that is one of the areas that I think is very important.

**The CHAIR** — You highlight there ‘low ethical threats’.

**Ms HORN** — Yes.

**The CHAIR** — One that comes to mind might be diabetes or heart disease.

**Prof. ROSENBERG** — Yes.

**Ms HORN** — Yes.

**The CHAIR** — Governments are sensitive to potential landmines in research and data. Going to my question, which is a follow-up from Marsha’s, you indicated that it is important to develop a whole-of-government licensing approach and it would also be useful to have a national approach rather than a state-by-state approach. We have seen a lot of evidence that Queensland appears to be leading in this area. Would you like to comment on whether we should be building on Queensland, whether in our report we should set new directions —

**Ms THOMSON** — Or whether Queensland is leading?

**The CHAIR** — I was not going to put it quite as bluntly as that. Whether we might like to set a slightly different course that would be more insightful and others might like to come on board, given they have done groundbreaking work to date?

**Prof. ROSENBERG** — Anne made the very sensible decision to move from Queensland to Victoria and to Deakin University.

**Ms HORN** — Groundbreaking.

**Prof. ROSENBERG** — So Anne is in the best position to — —

**Ms HORN** — I think a lot of work has been done in Queensland. I understand advice is being sought there from some key researchers up there who have assisted in that. There are the developments in Queensland, and I am interested to see the proportions of what they say they are making available openly and what may be — —

**Ms THOMSON** — Sorry, can I just butt in there, because what I would really like to know is not so much what they make available, because we make heaps available.

**Ms HORN** — Yes.

**Ms THOMSON** — What you have complained about is whether or not it is findable?

**Ms HORN** — That is right.

**Ms THOMSON** — And maybe it is a case of too much information and it is hard to source; the second is, how useable it is. If you have said the issue is the ability to find it and the usability of the data in the form that it is in, if you have got any information in relation to Queensland’s versus Victoria’s accessibility on that score that would be useful. Rather than get



caught in the term 'Creative Commons', the practicalities of access are what I guess I am really interested in.

**The CHAIR** — But in a whole-of-government approach?

**Ms THOMSON** — Yes, that is right; absolutely.

**The CHAIR** — You are suggesting we need to do this nationally? How do we do that national approach?

**Ms HORN** — In suggesting that I am recognising there are strong national positions being taken, particularly where we sit in the university sector, that we now have increasingly, obligations to make any research we undertake — whether it is with industry partners or government partners — openly available, to make the output, what is published, openly available, and it is not always so easy to get the published information from the Victorian Government.

**Ms THOMSON** — No, I accept that.

**Ms HORN** — Which is the easier part.

**Ms THOMSON** — I have tried to use our search engine.

**Ms HORN** — Yes. So we are increasingly under obligations — being driven by the review and the recommendations in the venturousaustralia report, the accessibility framework, and the funding rules of our ARC and NHMRC grants — to make whatever we do in the search, as far as possible, openly available, and work with our partners to do that as well. That is the national approach that we find ourselves — that is what we find ourselves in, being driven by the national government. I cannot draw a personal comparison between how good Queensland information and Victorian Government information is.

**Ms THOMSON** — I am not asking for a personal account, I am just wondering if there has been an assessment from a research perspective about trying to access this kind of information nationally and whether we are all on a par, or whether there are some easier ways of accessing some versus another.

**Ms HORN** — I am not aware.

All I can say is the suggestion was made to me that it was easier in Queensland and South Australia than in Victoria. But that is a third party experience.

**The CHAIR** — Perhaps we could follow that up; perhaps Vaughn and Yuki can send you a letter regarding that honing in on a national approach and honing in on state by state — what works well in some states over others and how on earth we are going to get those, that are not doing what one state might be that you see as groundbreaking, on board.

**Ms HORN** — Yes, okay.

**The CHAIR** — Our task is not to tell other state governments what to do, but to give a compelling reason why we have come up with a particular resolution or recommendation and they may like to come on board.

**Ms HORN** — Yes.

**Mr DAVIS** — Just clarifying this national perspective, what you are in effect saying is that your funders, both departmental and research, require information to be made public in the large majority of situations, and that drives you to also want to make things public, and I accept that completely. But that is not quite the same as the debates we are having also about whether the large variety of state data that is available there should on its own account be made available as

much as possible in its raw form, perhaps, or in metadata form. You are obviously supportive of that concept that as much as possible should be available.

**Ms HORN** — Yes.

**Mr DAVIS** — And the advantages for the state government, the state economy and the state research sector to do that.

**Ms HORN** — I do not see that it is a different argument for a national basis or a state basis.

**Mr DAVIS** — That is exactly my point. The two arguments may actually work in tandem.

**The CHAIR** — Have you got a question you would like to specifically ask?

**Mr DAVIS** — No, that was my main point. Although just picking up Evan's point about metadata, which I think is a central question for our inquiry about the balance of what you make available, I would have thought that as researchers you would see it would give the Victorian university research sector, if I can coin a phrase, a significant national and international advantage to have access to a large amount of data.

**Ms HORN** — Not necessarily. Certainly the researchers are very keen to have access to a large amount of data to pursue their research. As I said, they look keenly elsewhere at other countries where this is not an impediment — and they do see it as an impediment.

**The CHAIR** — Thank you, that has been very helpful. We will be forwarding you a copy of the Hansard transcript. You will have the opportunity to correct typographical errors but obviously not to change the substance of what you have said. We thank you very much for your assistance both in your presentation and your submission today and any follow-up work that you forward to us or we communicate to you.

**Witnesses withdrew.**