

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

Melbourne — 8 September 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

Witness

Ms L. O'Brien, Vice-Principal, Information, and Chief Information Officer, University of Melbourne.

The CHAIR — Welcome to Ms Linda O'Brien. The committee that you are appearing before today is an all-party parliamentary committee and is hearing evidence in the Inquiry into Improving Access to Public Sector Information and Data. You are very welcome. All evidence taken is protected by parliamentary privilege. Any comments you wish to make outside the hearing are not afforded such privilege. Could you please state your name and address, if you are attending in a private capacity or representing an organisation, and if you are representing an organisation, your position. Give your professional address, not your personal address.

Ms O'BRIEN — Linda O'Brien. I am representing the University of Melbourne. The address is easy: it is University of Melbourne, Melbourne 3010, and my position is Vice-Principal, Information, and CIO.

The CHAIR — All your evidence will be taken down by Hansard and will become public evidence once you have had the chance to check the draft by Hansard. You are free to make any corrections to typographical errors but not to substantial content errors. Over to you. We look forward to your submission. Thank you.

Ms O'BRIEN — The university has put forward a formal written submission so I will just speak briefly to that submission, if that is okay. I assume you may wish to ask questions at the end.

The CHAIR — Yes. Mr Crisp will ask questions and you have the opportunity to also have questions asked by Dr Koops as well.

Ms O'BRIEN — Thank you very much. The University of Melbourne has just been through a six-month process looking at scholarly information in the digital age. Many of the issues that we have canvassed within the broader university community and beyond are common to this particular inquiry. We engaged our academic community and the broader community, including external advisers from both the United States and the United Kingdom in this process, so it has been quite a thorough investigation of issues. As we know, the world is changing rapidly, but it does not appear that that will change the way information is made accessible and available and used. What we found is that there are real tensions at the moment between open access for everything or use of licensing to protect products or intellectual property; serving the public good and being financially and environmentally sustainable in our practices — again, similar tensions that you will see in this inquiry — centralising information, data, documents or distributing them into multiple repositories and working out what the best blend is between distributed and centralised systems; and creating your own search, cataloguing or other tools, and populating tools that exist nationally and internationally. There is not a single answer to these questions and they are complex questions for a university like the University of Melbourne where much of its research is international in nature and where we are partnering with universities and other organisations beyond national bounds, let alone state bounds.

What we believe will give us the sorts of opportunities that we need to exploit in the future is to really emphasise flexibility, so in any solutions that we are putting in place we make sure we adopt processes and procedures that enable rapid, tailored responses to emerging circumstances because they will keep changing. The word 'podcast' has been in the vocabulary for less than five years, which is fairly hard to imagine. Now three-quarters of the refereed international journals are in digital format and the predictions are that the scholarly text will be digital within 10 years as a normal way of offering it. Flexibility will be critical in anything we do. We need to be agile. We are proposing in our vision that our IT systems be set up as small, loosely joined pieces, not large enterprise systems that attempt to do everything, because that will allow us to move in and out of different products as the world changes and to plug into different systems. Open standards will be critical and we need to adopt international open, not proprietary, standards for describing and organising our information if we want to ensure its longevity and that it can be easily shared and reused. Collaboration will be key. It is clear that collaboration across the university's research

institutions, public and private sector, will be an important part of leveraging collective strengths and that will continue.

I guess they were our key findings in terms of how we are shaping our thinking about building our information environment into the future, and they mean as a result that we have put a lot of emphasis on things like appropriate identity and access systems, and on working in partnership with other key agencies to put our information in the best place to leverage benefit for the university.

We have suggested in our submission a range of possible pilot projects, if there was an interest in pursuing some pilots, where our researchers have expressed interest in the use of public data to further their research outcomes, and they are quite varied as you will see from the document. For example, looking at medical errors for the purpose of quantifying risk would involve gathering information from a variety of sources to manage and manipulate that data and advance our research understanding and thereby inform practice. Some other things that are, I think, important issues for the government at the moment include water and power usage statistics and how institutions are starting to respond to the challenges of sustainability issues. Another that was suggested was looking at the liquor licensing laws and their effect on inner-city residences and trying to garner information from multiple government sources that would allow us to understand what is happening.

With those sorts of projects in mind and going back to those principles, it would therefore suggest, in terms of possible recommendations, that open standards are adopted for data formats and for metadata so that it is very easy to be able to access information in a way that allows those sorts of research and teaching and learning outcomes to occur. It would also assist the authentication systems that are put in place around access to that data and use of the data to comply with the Australian access frameworks that are currently being developed, which will provide the depth of identity and access management required to minimise the risk of information becoming inaccessible over time through using appropriate standards, and also allow that data to be used in new ways around authentication and access.

We also suggest as a general principle that publicly funded information should be used for public benefit and to just use that as a given in terms of the approach that is taken. Again this was an issue tackled within the university; clearly there are times when information may be of commercial value but by and large the bulk of the sort of scholarly output that our researchers develop, whether it is the research data itself or the research outcomes from that data, is of immense value beyond just a commercialisation opportunity. There will be times when that needs to be protected for those reasons but more often than not public benefit can come from sharing that underlying data and also the research outcomes. Certainly that is the principle we are employing with respect to our publishing in the future.

We also suggest that there be a single set of policies and procedures, and again even within the university we are attempting to start to get this sort of approach in the way in which we manage our information and determine what information is released and under what conditions. Certainly it would make our engagement with Victorian Government agencies much simpler if there were common sets of policies and procedures that were applied consistently across the departments and agencies.

It would be critical that a suitable governance structure be put in place to ensure ethical management and use of the information being released, and we would propose that that would involve government, business, professional users and consumers. It is through such a governance body that matters requiring confidentiality and processes and decisions should be transparent and open to public scrutiny, but to have an open process by which decisions are taken. Within the discussion paper document you canvass issues around Creative Commons. A governance structure would satisfy this, probably as a way of licensing a large part of the government's information licensing needs, but you may also wish to look at your own licensing framework. I

have actually met the creator of the Creative Commons licensing. It is evolving and getting a lot of purchase throughout the research and education sectors. It would be a matter as to whether you felt it was easier to adopt existing conditions and licensing that are already out in the public domain or whether you wanted your own licensing framework, with some template licences, that could easily be applied to different types of data and for different uses and relationships between creators, custodians and end users of licensed information.

I think those five points are the main recommendations that we will be making. There is a huge opportunity to leverage the investment in public information for public benefit, and universities such as Melbourne could clearly assist in taking that to the next stage. We would be very keen to work with the state government. We already have a very strong relationship, particularly through Multimedia Victoria and initiatives such as the Victorian eResearch Strategic Initiative, so we would be keen to continue. Life science is an obvious area, with current investments proposed around life sciences computation.

Mr CRISP — All right, thank you. I will kick off with one that comes to mind from previous evidence. You talked about how you are going to hold and manage the information — dispersal of the information versus centralising.

Ms O'BRIEN — Yes.

Mr CRISP — And this has come up before, because you have got territorial managers of information who would be very reluctant to give it up. Also, equally, while appearing to provide easy access to information if it is centralised, there are complications that come with that. I would like you to talk a little more about how Melbourne University went through that process of managing the people, the territorial wars and the outcomes to arrive at deciding to leave it dispersed but having access to it.

Ms O'BRIEN — Yes. The reality is, if I take our researchers in physics, for example, through the Hadron collider, which went live last weekend, they will be generating over the next few years petabytes of research data. We have individual research groups that will generate huge volumes of data. The bulk of that data will be actually hosted at CERN in Europe but we will also host data within Australia, as the Australian node for the Hadron collider experiments, at Melbourne. Inevitably we are in a distributed environment. Before we even begin the world we inhabit will be distributed. If I take our law researchers, with their research outputs, their publications, they put them into an archive, a social sciences research network archive, which is hosted in Europe. That data is then used to rank the institutions based on the citation impact and ranking — so the amount of use that data is given is used to rank universities and therefore provide some indicator of quality. It makes much more sense for them to put that research output in that database than for us to host it in Melbourne because it adds benefit. Pubmed, in medical sciences, which is published out of the US, is the obvious example. It is the place to put your medical research output in terms of maximising international impact, and it is the place that even researchers within Australia know to go to look for that sort of material.

Inevitably it was a distributed environment. If we translate that back into local practice, what we will be doing is offering to provide some of that core underlying infrastructure so that we can maximise our investment by having large storage, but the way in which technology now works and the way in which you can virtualise your environments means that where it is stored is no longer an issue. It can appear as though it is at Melbourne even though it is somewhere else. What we will do is try and maximise the value by having some shared storage infrastructure but allow people to virtualise at the next layer, and allow them to place their content where it best sits, but then through metadata be able to show that as a complete suite of information even though it is hosted on other sites.

It is pretty clear that the horse has already bolted; it is a distributed environment, so let us leverage what technology now makes possible by making sure we have got the standards in place and the

frameworks in place. Where it makes sense to aggregate to get more bang for the buck, like commoditised storage, we will do so, but even that is a matter of time. I think within the next 5 to 10 years we will see that the commercial providers will have offerings there that make it more sensible that they host data on your behalf. We will still need, as a long-lived institution, to be the custodian of that information and ensure that it is preserved in perpetuity, but other providers will start to provide those solutions.

Mr CRISP — We have been focusing on the three — copyright, Creative Commons or open source — and you mentioned that you could do your own licensing. When it comes to government, I have the view that if it does not come out of a box, do not touch it.

Ms O'BRIEN — I can understand that.

Mr CRISP — But I was interested, because that is a section of the debate that I have not heard yet — that you could write your own licensing stuff if you so wished. I guess I am just feeling around now, because that is new, as I said, and my instincts are if it is not in a box — —

Ms O'BRIEN — Yes. I guess it depends whether you want to do an adaptation of Creative Commons, which others are doing; it is not an uncommon thing. They take the Creative Commons framework and adapt it for their particular need so that they have a suite of standard licensing and template conditions that different agencies can pick between in terms of the way that they wish to manage and make their information available. It is a question of whether you want more nuanced approaches. There are moves afoot. You would probably know better than I what is happening at a national level, but certainly from a research perspective things are occurring at the Australian National Data Service on thinking about a national framework around Australian research data, which will of necessity bring in all the research organisations which are government instrumentalities of some kind. That group will be developing frameworks and standards and approaches which clearly the Victorian Government could piggyback on. It depends on the extent to which you want something that is more locally nuanced, or — again, you would know better than I — whether other states are grappling with similar issues and there would be merit in a kind of federated approach to thinking through those issues around government data.

There are many different models out there. It is really a matter of what would work best. Certainly within our strategy we have said if we can find an existing example elsewhere, we will adapt it and adopt it for our own use. We will try and stay with open standards, open framework — so we are using Creative Commons within the university, and that is our preferred model, not to invent our own. It is a question of whether there are particular circumstances that you would like to be able to nuance in a different way, because Creative Commons has come out of the academic community; it is about furthering research and advancing learning. It has that flavour to its approach. It is whether you want something that is a bit more nuanced.

Mr CRISP — In your submission you talked about some obstacles that need to be overcome for open access to be viable. Do you want to run through some of those for us in examples so we can get a feel for where we are heading?

Ms O'BRIEN — Okay. There are many. I guess some of them are around the legislative challenges. The proponent of Creative Commons argues that copyright is dead, that it is increasingly difficult — again, this plays out differently in the case of government information, but certainly from an academic perspective — because information is increasingly created collaboratively and the scholarly outputs are creative endeavour; the notion of who owns the right to that copy is becoming more and more complex, and it is whether copyright will last the distance in terms of what is happening with publishing. We see it even with our students who are creating content in collaboration with our lecturers: who owns that content and how is that managed? The intellectual property rights are often clear, but who owns the copyright is a much murkier issue. That may not be as much an issue in the case of government data, but certainly one of the

obstacles for us around open access is concern about opening up that content and what that might do.

There is the tension around commercialisation that I mentioned. The reality is that globally very few universities make very much money out of commercialising their intellectual artwork. The real public benefit comes from making that accessible. But, that said, many of our academics still feel they wish to hang on to their content and not make it openly accessible. Our educators, for example, feel quite fraught because the research ranking system is heavily biased towards citation impact now and publication in refereed journals, but they actually want to get the research output into the hands of the practitioners to improve teaching practice. There is this real tension in the model at the moment about getting it as open as possible so it can get to those who can make best use of that research, yet furthering the research profile of the university through the ranking system. There are a lot of tensions in the model at the moment, some of which are less relevant in terms of the sort of content you would need to make available. The other is just the whole issue of sustainability — the sorts of formats and ways in which you would ensure that you do not end up with something that cannot be sustained in the long term from a technical or a data perspective.

Mr CRISP — The next area we want to explore is how the access to government data is now difficult to access. We are looking for good and bad here. You described some of it even as restricted. You used the medical data example earlier, but also hopefully there is also some good and bad. Are there any other areas of government that are good at getting this data out, as well as the ones that are not so good?

Ms O'BRIEN — I probably would not be as well placed to talk specifically about Victorian Government data and what works and what does not. That is mainly because at the moment most of our academics individually try to approach whoever it is and access content. I guess that is one of the issues at the moment. Those individuals undertaking research will make direct contact. There probably is not a framework in place to make that easy, and it probably means you are being overwhelmed by individuals going to individual officers about particular aspects of information. Some really good practice examples that I am aware of are groups like the Australian Bureau of Statistics, which has been for a long time very concerned with how it makes its content accessible to the academic community. There are good examples out there. At a Victorian level I am really not well placed to give you specific examples, but we would be happy to follow up if that would be useful.

Mr CRISP — Thank you. I think you have covered the question 'Is the university system part of the public sector or is it separate?'. You are very keen to be separate. You are obviously developing your own model. You are planning to have your own model for information access or dissemination, separate to the public one. We picked that up, which I suppose puts that to bed. I do not know whether you want to add any more to that on whether that has pluses or minuses. The university obviously has made its decision.

Ms O'BRIEN — Yes. I think it is about the fact that we operate not only at a state level but at a national level and an international level. We need to be really following international best practice in terms of open access in terms of data sharing. I think it would not necessarily sit comfortably therefore to be within some state-based framework. I think we adhere to the same principles we would like to think the state government would adhere to in terms of starting with 'open is best' first, in following open standards and ensuring good practices in terms of sustainability — a lot of the things that we would like to see would be a government approach. I think the principles will be the same. It would only vary in the practice in as much as we will be taking our lead from international initiatives. Groups like CERN in the physics area or, in the medical research area, the US medical research organisations will be really ensuring that our academics can operate within those international frameworks in the most effective way, because that is increasingly the way in which research will be done.

Mr CRISP — I guess last of all it comes back to standards for metadata. If we have these different systems operating, then it comes back to the key that we are going to have to get the metadata level right.

Ms O'BRIEN — That is right.

Mr CRISP — Where is the university up to with regard to that? What would it make as its contribution and what sorts of standards in the metadata system would it be looking for?

Ms O'BRIEN — Metadata is a fascinating one because what we have seen at a national level is that there has been an interest in allowing the disciplines to determine the metadata frameworks that work within their discipline. What I have been arguing to anyone who will listen is that we need some standardisation across those frameworks, because it is the cross-disciplinary, multi-disciplinary research where there will be real wins in the future. There is a place for acknowledging discipline differences and nuancing a framework but within a national and international framework. At the moment, what has happened is that most discipline communities have developed frameworks that are starting to get a fair amount of buy-in internationally in the research community around a discipline cluster. What we have said at Melbourne University is that what we want to do is ensure that there is something layered over the top of that that provides a consistency in our metadata to allow that cross disciplinary search to occur. This is something the Australian National Data Service will also be grappling with at a national level: what sort of metadata schema? Melbourne will be one of the key players in helping to try and influence that at a national level. There will be a need for a common approach but then with acknowledgement that there are discipline differences that would require different schema. That is the approach we are taking at the moment. We are trying to keep tabs on what is happening at an international level across the disciplines. So we are engaged with UC Berkeley and a number of other universities in the US and Europe in the humanities at the moment, looking at humanities metadata and how we might make sure that what we are doing is all aligned. I think we will be looking at those international trends but trying to make sure we are layering metadata that gives us a common framework as well. Hopefully the Australian National Data Service will help us with that.

Mr CRISP — That was good. Thank you very much for that.

Ms O'BRIEN — I hope that was helpful.

Mr CRISP — It is. We are putting this together. You will receive a copy of the transcript in about a fortnight. Typing errors may be corrected but not matters of substance.

Ms O'BRIEN — Certainly.

Mr CRISP — Again, thank you very much for your evidence today. I think it is an exciting area that we are putting our toes in.

Ms O'BRIEN — It is wonderful to see that the state government is thinking about this, because I think there are huge opportunities here. There are plenty of challenges but real opportunities to do some very clever things that will have broad benefit.

Mr CRISP — Thank you very much.

Witness withdrew.