

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

Mr C.G. Obst, Regional Director, Victoria, Australian Bureau of Statistics.

Mr CRISP — Welcome, Carl. There are some formalities we have to attend to first. If you have been before a parliamentary committee before, you will probably understand this. This is the Economic Development and Infrastructure Development Committee. We thank you for your submission and welcome you here today. It is an all-party parliamentary committee that is hearing evidence today on an Inquiry into Improving Access to Victorian Public Sector Information and Data. All evidence you give today is protected by parliamentary privilege; comments you make outside the hearing are not afforded such privilege. To help us get it on the record, I need you to state your full name and business address, and if you are attending in a private capacity or representing an organisation, which I can see you are.

Mr OBST — My full name is Carl Gordon Obst, and I am representing the Australian Bureau of Statistics, based at 485 Latrobe Street in Melbourne.

Mr CRISP — Thank you very much. The evidence that you give today will be taken down and in time will become public evidence. I am going to invite you to make a verbal submission to support your written submission, and then we will have some questions at the end.

Mr OBST — Just a few introductory comments: it is very clear to us that there is a changing demand for the type of information people are looking for, and the detail with which that information is being sought is changing over time. It is also true that while it can seem like a very recent phenomenon, demand for information is something which has been around for a long time, but at the same time demand has been changing for a long period of time.

Part of that demand for information that we have been seeing growing in recent years has been driven by technological developments that have allowed access to information and asking questions of information to a greater degree. But a lot of it has also been driven by changing societal factors and an increasing demand by society to find out about the world in which they live in and the things that affect them across environmental, economic and social factors, in particular a demand for information with a very fine level of detail around communities. There is a very strong movement in a number of places to try to get access to that type of information.

I think there are two elements to that. One is to try to get an idea of large, changing trends at that fine level of detail, with broad questions about whether people are better off or worse off, and the general notion of looking at progress and how that can be assessed beyond GDP and measures of economic performance. But then there is an increasing demand we are finding, perhaps in response to those questions of looking for detailed data, to try to really tease out how you might actually respond to those broad questions.

What that focus on that detail means is that it is often only information that is held by the public sector which is able to get at those real questions of detail. So from an economic perspective you might be looking at information provided to the Australian Taxation Office on every business. How can you start looking at that information in a way which allows you to tell stories about what is going on in the economy, that you cannot get if you look at larger, economic surveys? Because they are designed to give you an all-Australia picture, or along those lines. It is the same sort of thing around social data and environmental data; people are looking for really detailed information. That places some significant challenges on both the ability to get that type of information out and also to protect the privacy of people whose information is contained within those systems.

From an ABS point of view, we are partly fascinated by this whole information scene and how we can play a role. We have an independent, legislated role to be looking at elements around information. But we have been trying to look at these types of questions for over 100 years, and very much looking at things from an end-to-end perspective about how that data is put into the system, how it is managed through the system, and then how you tell stories about it and disseminate it at the other end. Our recent thinking about the ABS role — I guess for a long time there was the perception of it sending out forms and getting people to fill them in and getting complaints about forms, as I tended to do. Having a bit of a rural background, I used to find myself going out and getting a heap of things from farmers at different places and complaining about forms. We are trying to see that we are placing ourselves now as, rather than just being a survey collector, playing a much more strategic role around information.

We have tried to outline in the submission four areas where we think we need to do that. One is about getting a general recognition of the value of statistical information, seeing that in a sense knowledge is an asset, that type of

view. One is around building the capability of people to actually use the information and interpret it, both in terms of professional analysts but also in terms of what we call statistical literacy of the general population.

The third angle is about looking at content and making sure we have got information that is both detailed and also relevant to the types of questions people want answers to. The final one is about looking at infrastructure — that is, things like standards around information definitions, but also looking at the procedures and how you might best practise in collecting data and managing it and then getting it out the other end.

Largely we see this as an issue which to take forward effectively is not something ABS can do by itself; we need to have partnerships with particularly other public sector agencies, both commonwealth and state and territory. That whole direction of developing strategic partnerships around information is a key aspect of where we would like to keep pushing our strategic directions. One thing that I guess we tried to get across in the paper is that we are very clear about the benefits of information and the potential that access to information can have for driving both broader public good and individual private benefit. But there are a number of complexities and difficulties in doing that.

One is, as I mentioned, around privacy and trying to make sure that there is a high degree of both perception and reality of people believing that their information is safe when passed on to the government. Now I think there are probably different schools of thought about how well that is done in terms of the perception. Some people might perceive that all government departments must be sharing information, and I guess there are different ranges around that. I guess, very much on the perception side, we have seen numerous examples in different countries of where a loss of trust in the public sector and in the ability to protect information has really started to rebound quite strongly in terms of the willingness of people to accept information from the public sector and to see that it can be used effectively for policy making. There really is a strong requirement on public sector agencies to protect that privacy, to retain that public trust which feeds itself.

The other complexity we see is the importance of linking all of these end-to-end developments, so we need to think very carefully about how data feeds into the system, how it is managed through it, and how it is put out at the other end. The discussion paper picked up on a number of licensing issues and use-of-technology issues. We would see those as one part of the thing that we need to deal with, but there are number of other things that need to be considered as well if we are going to allow this improved access to public sector information.

I guess the final comment to make is that we are very keen to push forward for a national approach as far as possible, and a consistent approach. The biggest benefits are likely to be gained where we are operating in an environment in which Victorian data can fit within a broader scheme of data sharing and data access across Australia. That is probably all I need to say in terms of introduction. But I would be very happy to answer any questions associated with the submission.

Mr CRISP — I am going to ask you to comment about when it is free or charged and when you have cost recovery. The ABS provides a basic information set free of charge, but there are circumstances when you do then charge for data. You have a commercial pricing policy for services. Where is ABS sitting with all of that now? Which direction are you going? Are you still maintaining those commercial links for cost recovery, but as you make more available you get less cost recovery? How does that fit within the government? Or do you still have some good selling products?

Mr OBST — I think the starting point is that we are not trying to put a price on the information itself. It is not that we are saying that the data does not have a value; it is just that we are not trying to put a price on that. So when we charge for things it is not the case that that information is inherently more valuable than information we are providing free of charge; that is a starting point. In terms of making that decision about whether we charge for something, we really have to make a judgement in a sense of what our funding allows us, that has been appropriated from government, and is essentially requiring of it. There is a sort of basic information set that tends to grow over time as we get requests. And if we get a number of requests for exactly the same thing, then we might start thinking about whether or not we should be presenting data in a slightly different form. But normally that will be within the remit agreement of the set of surveys or collections that we are undertaking within a particular period of time. Generally we will find that where we are charging for services it is because we are manipulating the data in some way, which would not be a view that someone would normally see. Sometimes that means we have to undertake extra work in order to, what we call, confidentialise the data. Whenever we release data that is free on the Web, we make sure that no individual or business can be identified within that data. When people request an

alternative view — perhaps a more detailed view, or simply just a different way of putting different things together — it may mean that we have to undertake additional work in order to see whether or not that view of the data creates a potential for identifying businesses or individuals. So really what we are charging for then is the extra work that is required.

Mr CRISP — Value adding.

Mr OBST — It is value added. In terms of how that is changing over time, I think in actual fact we are probably finding that we may be getting more client servicing. I have not got the particular information on this, but it is not that we have had a drop-off in work around that. In actual fact, if anything, a demand is created because you are presenting a lot of data, and people say, 'Oh, if I've got that much, then perhaps I can just take a different view of that', and they are willing to pay a little bit extra to get that.

The other thing that we are finding is that we are moving probably more into also providing value-added services beyond the data — so providing technical support for people to put together a survey of their own, if they are inclined to do that — particularly working with the Victorian Government and other governments around Australia to develop their statistical capability, training services, those types of things, which we are doing basically on a fee-for-service-type basis. We have attempted at times to put a notion of core statistics, or the fundamental statistics, or the ones that we would not do without, and those are the ones that would be free. It is not something that can or should stay static, in a sense. In some cases it might be what we really should be measuring, but we are not at the moment. How you define that boundary has always been quite tricky. It is a bit of a fluid thing.

Mr CRISP — The next one we want to look at is how you have managed with the National Statistical Service. What would you recommend by way of governance mechanisms and infrastructure frameworks to facilitate the improved access to and the reuse of Victorian government information? You guys have crossed a bridge. Any recommendations for us on the mechanisms you used to get there?

Mr OBST — I think the question of whether or not we have crossed the bridge, certainly we have crossed the bridge internally in that we are pushing that direction more and more. It has been and will continue to be for a little while a little bit opportunistic in trying to find opportunities to find that partnership. The NSS is essentially a series of partnerships with other government agencies, commonwealth and state and territory. As we head down this line and try to push for greater recognition of the value of statistics and the way in which they can be developed and used, we are starting to tease out some of those governance issues. I do not think we have a solid framework in place in the sense that while we certainly do govern our ABS operations using a strong governance framework — we have the Australian Statistics Advisory Council, which is a body which sit above the ABS operations — there is not a body that sits above overall NSS directions per se.

There is a model in Queensland, where they have a Queensland statistician who has much more of a controlling role, so to speak, around the Queensland statistics, but it is not necessarily clear that that is a model that will be perfect either. I think the challenge is to try to work within a number of existing governance-type arrangements to see how information can be shared and opened up. The ABS is looking to play a leadership role. That is certainly true. In which case the way in which the ABS is governed is particularly important, but the way in which we interact with others at this stage is something that is still developing. It is probably too early to say that we have a governance framework in place. I think there is certainly some potential within the Victorian Government. From my exposure to it in the public service over the last six or seven months since I have come here, it is certainly a whole-of-government flavour that is much stronger here than in other states. But again developing that presence around statistical information can be a little bit tricky because the benefits do tend to be quite diffuse; they are not ones that accrue to particular departments or particular groups of individuals.

Mr CRISP — The next one we are looking at is what we call the Queensland project, which is the collaborative work done at the Queensland Office of Economic and Statistical Research and QUT to develop a licence manager. Can you outline the key components of the project, including how that licence framework will enhance the capabilities of Creative Commons in that model, which is very much where Queensland seems to be working with Creative Commons? Will it be implemented, and how is it going to work at the end? Can you see that far ahead yet?

Mr OBST — What I might do is perhaps get someone to be in more direct contact in order to give some of the details of that, because I think it probably needs an expert to talk to it rather than me. Our feeling at the

moment is that the Creative Commons framework is a good starting point but does not solve all of the issues. There is the potential to be able to have some technologies which allow you to deal with more than the Creative Commons licensing situations. The idea is to come up with something which is quite a flexible tool for dealing with a large range of licensing situations. Where we are at the moment is that it is still in the development and piloting phase in trying to put the software together. They have had quite a number of challenges in trying to do that. My understanding is that they are quite hopeful of getting something done towards the end of this year, at which point it will more of an assessment of exactly how that will be taken forward.

There is this separate project of looking at how in fact we can use those types of licensing arrangements on the ABS website as well. My understanding is that that is linked to the way in which search engines are heading. Google is trying to look at giving some priority to those types of data sets that have been licensed and recognising that sort of thing. There certainly is a lot of consideration going into that at the moment, but I think it is too early to tell exactly how that will play out. In a sense it is early days in terms of saying we have got a solution around it, but the directions at the moment seem to be very positive in terms of thinking that that is something that can work. If it would be of assistance, I could get someone who is really up-to-date. Because it is something that is going to be changing fairly rapidly over the next months, perhaps it would be worth keeping in contact with them in terms of finalising where things are up to for your report.

Mr CRISP — I am very happy to let that one go on notice. The next one I want to look at is the nationally consistent approach. We are looking at public sector information in Victoria. I am interested in your thoughts on having a nationally consistent approach, and the benefits of a national approach. Does the ABS view Creative Commons as a viable option for a national approach?

Mr OBST — In terms of the last question, the viability of Creative Commons will come out of what we work on and finalise over the next few months. More generally, we are certainly very keen to try to see whether or not we can find those solutions that would work more generally, notwithstanding the fact that each state has a different starting set and getting into understanding those types of things. The other benefit I have come across with Creative Commons is the international link as well. It would be a pity to design something that was nationally consistent but was not consistent with the rest of the world. There is that dimension around Creative Commons which has good potential as well.

Mr CRISP — The last question I have looks at the privacy component, which you said is so important in trust. As to the level of having a national approach, in supporting other jurisdictions we always worry about how the states and the commonwealth are interacting with each other and whether we should be doing our own thing or whether we should be working collaboratively. I am interested as to whether you have had any discussions on this and some of those outcomes around building those privacy structures nationally.

Mr OBST — I think our interaction in this space is growing, but the outcomes are less than clear. We have had input into various works of the Australian Law Reform Commission. We have been looking at issues around privacy, and we are starting a review on secrecy as well at the moment. The issue has also become more complex in recent years through the demand for and desire to link information from different data sets, which complicates the picture somewhat more.

I think we have been historically in the mode of looking at a particular data set and making sure the information within that data set is protected. What we are talking about increasingly is this sharing of information or joining information together across data sets, which creates all sorts of additional tensions. My reading of the information privacy legislation within Victoria and our consideration of that suggests that we are sort of basically heading in the same direction, or our understanding of what would be required is basically on par. I think the issue here is probably going to be one about bringing along the constituents as much as anything. If different states are at different levels of understanding of where their constituents are in terms of acceptance of sharing information, then it is going to make it a different legislative environment.

I was looking for, I think, the answer to the question you have got in trying to put together this submission about having a clear statement of exactly what the ABS would put into a piece of privacy legislation. I think we have still got to be looking at that work a little bit more. A lot of the ABS's own legislation about what it needs to do to protect the privacy of its information is very clear, and we are quite comfortable in seeing where that is placed. Our ambitions to try and lead a statistical service are pushing us towards then trying to consider legislation in other areas. Every now and then we run into that question, for example, around terrorism legislation which was

introduced a number of years ago, about whether or not that required the ABS to give up information in response to terrorist threats and the roles of the ABS around that — the sorts of examples where we had to consider where the ABS legislation sits within that. But the broader context of what privacy might be relevant in different states and nationally is something we are still working towards, which is probably not necessarily helpful. Perhaps it goes to that point that particularly in this area we are looking for partnerships and working together to try and see what would actually work in this type of space.

Mr CRISP — I want to pick up a point there too, because you have given a very good indication of where we are with the privacy staff. You mentioned an inquiry into secrecy. Is that within ABS?

Mr OBST — No, that is the Australian Law Reform Commission, which is starting a review, I understand, looking at secrecy, which is a slightly different issue from confidentiality and the protection of information. It is more around what information — I might have to think about that one as well. The secrecy one is what information needs to be protected over a long period of time, whereas the concept of confidentiality at some point in time does not become as heightened. So secrecy is something you need to protect forever and a day, whereas confidentiality can be something that you would protect but the circumstances may change over time such that that information is not necessarily as confidential. They are very related concepts, but they are not quite the same in legislative terms.

Mr CRISP — That is obviously of interest to the ABS. You would only be interested if there are some data sets — you have an internal debate about privacy, confidentiality and secrecy, so there are data sets going forward that need to be considered for secrecy obviously, which interests me.

Mr OBST — Yes, it does have a strong bearing on what we do and making sure that we understand our legislation in terms of its current interpretation and what the potential is. I think it is a dynamic. The social acceptance of information and the need for it and balancing that sort of protection of individuals versus some collective good that could come from increasing access to information is a careful one to keep in balance. It is easy to go one way or the other in terms of how your legislation might move forward. But I guess our view is that there is a public good that comes from information, which is hard to put a value on, but nonetheless is quite important to recognise, and I think community attitudes to the extent of that public good can change over time. So I guess it is a question of continuing to monitor that and then seeing how that is balanced against any rights of the individual. I have not been involved in it, but I understand there is a charter of human rights in Victoria, and I think how that interacts with the notion of individual privacy is quite a fascinating legal question. I am not a lawyer, but there are some fascinating questions there about the public and collective good versus individual rights around information. It is quite a fascinating area.

Dr KOOPS — Could you give us a brief description of how the ABS organises its data? Do you have custodians of different data sets within the ABS or is it all brought into a central point? I am really going to the issue of metadata and how you coordinate all the data and work out who has got what, or whether in fact it is not an issue for the ABS.

Mr OBST — I would probably start with saying it is an issue for the ABS, one that we are continuing to learn about around the importance of metadata and descriptive information. In terms of the current environment we are a mix of both in the sense that each individual collection or survey area will have a set of systems and definitions and processes which it will use to create a set of outputs. So there will be a range of either survey information or administrative information that feeds into the process. It is managed through a variety of different things — technologies and systems — to produce a set of outputs that will go onto the Web or will be used then to create alternative views for users.

I would not want to hazard a guess at how many of those there are, but all of those then get fed into a central information warehouse, which forms the basis for the information that then gets sent out onto the Web. That information warehouse has a set of metadata or descriptive fields to try and cover off all of the pieces of information that feed into that. So in that sense we are on top of that metadata and how important that is in trying to drive different statistical products, whether they are spreadsheets or publications or those types of things. We are continuing to try and improve the range of products we have got coming out the other end in different formats that are available now, but again all feeding off that central metadata. What we have not got as good as we would like is the links between the metadata which is sitting on the output side in that dissemination and the metadata that is feeding in the various bits that do that. So there is not necessarily a lovely consistency between how we describe

data in the preparatory stage, so to speak, and how we are describing data at the other end. So there is certainly a lot of potential there.

Dr KOOPS — I am asking because even though you are describing some of the problems you have integrating those sets of metadata, I am assuming that you have a far more mature approach to your metadata than perhaps other governments' departments might have and whether there is a core set of standards or a core of lessons from the metadata that you could perhaps inform departments about so that they can — because the Victorian Government probably does not have a very developed metadata system for its data as against —

Mr OBST — Not as a whole, no. My general impression is that this is really a question of understanding the need to organise information and the costs involved in doing it. I think it is certainly the case that within the ABS we recognise the need for it and the potential for it. I think we are starting to see more broadly as we go along that once we have organised that patch, we could start to organise a better patch and then you start looking out and looking at the use of classifications and standards beyond the ABS and trying to share those types of practices. It is often difficult, but given that our job is information and statistics, then one would hope that would be the case — that we would be a step ahead.

My experience shows that statistical efforts in most departments are fairly small and somewhat fragmented. They try and put together a set of information, but there is not necessarily a lot of resource there to try and put it together and build it up over time. People are able to use a spreadsheet fairly quickly to get information together and present it and get it ready for someone to use, but it is often done on a one-off type of basis. You really only start to value the metadata when you are trying to do something a repeated number of times. If you are doing a one-off, a spreadsheet is probably the most cost-effective way of doing it, but if you are the person who has to pick up someone else's spreadsheet then you will probably start your own. The metadata comes as being of value as an investment-type of infrastructure where you are trying to create the same thing over a period of time, but getting that recognition of the value of doing that is something that we are working towards. One of the things that we are trying to build within the Victorian Government at the moment, as the ABS, is to try and get that recognition of the importance of that.

One thing we have also noticed is that with quite a number of areas within the Victorian Government they will have a very good understanding of why it is they want the information and what purpose it can be put to and how you would answer the policy question: why would it be good to do this?. It is that sort of policy link. You might also have a broad range of indicators that are available to inform you of that particular area, be it children and youth policy or education policy, but what is often trickier is to try and understand whether or not the way in which that indicator has been measured and the definitions behind it are appropriate for the concept that you want to try and measure. That gap between the measurement concept and the policy concept, as it were, is one that people find harder to appreciate. If you are not into enjoying the measurement-type of space, you can spend a lot of time at the policy end and then get to the measurement end and say, 'We need an indicator for that. Right! That sounds like the most appropriate thing'. It may be the only thing that is available and the best that can be done, but understanding the limitations of that and whether or not it is then heading you in the wrong direction is another piece of work that needs to be done if you are going to get some ideal sense of metadata. Without that link, the need for organising the information becomes somewhat redundant. If you can do all of your really good policy thinking and then just take any indicator into it and satisfy yourself, then organising all the information is not necessarily all that beneficial. You need to organise the information, but also link it to why in actual fact you are doing the organising in the first place.

It is bringing those two areas together which is at the heart of the statistical service. It is not much good for us to just go out there and say, 'You should organise your information because you will be a lot better off'. All that we would manage to do is organise a lot of data sets which are not necessarily related to anything that anyone wants to answer. That is that issue of maintaining relevance in your data set, but recognising the potential of the organisation is the tricky one to do, and one that necessarily takes a long time, because it is not something that can happen overnight. It is just a matter of pushing. That is not to say it would take 10 or 15 years; it is a bit of a mindset as well that needs to be adjoining from both sides, of both the people who like organising information and the people who want to use it, and drawing them together.

Dr KOOPS — In terms of a standard, though — a common metadata language across the ABS — is the Australian standard that they are talking about adequate?

Mr OBST — That is a question I will take on notice, because there is certainly a lot of work happening around metadata. My experience is that the work we are doing on the dissemination side is very heavily focused on trying to implement metadata standards. Whether that is sufficient or not, I am not sure, whereas bringing those standards back into our core operating systems and processing systems is not something that has happened to my knowledge, although I think we would be certainly wanting to head in that direction.

Dr KOOPS — Thank you.

Mr CRISP — Thank you, Carl. You will receive a copy of the transcript in about a fortnight. Typing errors may be corrected, but not matters of substance. Could you have a look at that, and then in due course it will appear on the Web. Thank you very much for your attendance here today. It is very much appreciated.

Mr OBST — Thank you very much. We will be in touch with additional information.

Mr CRISP — The Inquiry will work its way on. Thank you for taking that question on notice, because it was a difficult one.

Mr OBST — That is all right.

Committee adjourned.