ECONOMIC DEVELOPMENT AND INFRASTRUCTURE COMMITTEE

Inquiry into Improving Access to Victorian Public Sector Information and Data

Melbourne — 30 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 J. Wilbanks, Vice-President, Science, Creative Commons.

The CHAIR — I welcome Mr John Wilbanks, Vice-President for Science, Creative Commons, to this all-party parliamentary committee. Today we are receiving evidence on the Inquiry into Improving Access to Victorian Public Sector Information and Data. We will be providing you with a transcript of today. As part of the inquiry, people always state their name; if they are appearing in a business capacity, their business address; and their capacity within that business. Would you like to start with that, please, Mr Wilbanks?

Mr WILBANKS — My name is John Wilbanks. I am the Vice-President for Science at Creative Commons. We are hosted at the MIT Computer Science and Artificial Intelligence Laboratory at 32 Vassar Street, Cambridge, Massachusetts, United States.

The CHAIR — Thank you, and a special welcome, given the distance you have travelled.

Mr WILBANKS — It is only from the Sebel city hotel, so it is not that bad today.

The CHAIR — Could you perhaps begin by giving us a presentation, and then we will go to questions.

Mr WILBANKS — Do you want me to go ahead and read this?

The CHAIR — If you want to read it or if you want to take dot points from it, whatever you wish. If you wish this to be on the internet as your Hansard transcript evidence, then read it.

Mr WILBANKS — Okay, I will read it. I would like to begin by thanking the committee for the honour of presenting today. Both Australia as a nation and Victoria as a state deserve great credit for the attention and care paid to the issue of access to public sector information and its relation to innovation. I will speak today with an attention to science, which is the area where I work and the area I know best.

Global society faces a clear set of critical challenges: to eradicate poverty and malnutrition, increase our work against disease in support of human health, respond to climate change and manage dwindling natural resources. Science is the key to meeting these challenges, and the combination of modern scientific research and information networks promises to create an explosion of value, both in our understanding of all areas of science and in the development of policy and medicines to improve human health and the quality of life worldwide.

Unfortunately we have yet to see this explosion happen. Although web technology has brought tremendous innovation gains for commerce, it has so far failed to do the same for science. While the elements exist for an effective legal and technical infrastructure to spur innovation and discovery, we have not connected the dots. Scientists face multiple barriers to finding and using one another's research, datasets and materials. Rather than building on prior results, they are often left to duplicate work and follow blind alleys. As they miss opportunities for collaboration and discovery, the pace of translating basic research to drugs and other public goods remains static.

It does not have to be this way. We see in the consumer web new developments like Wikipedia, a collaboratively built and edited encyclopedia that rivals the expensive traditional alternatives such as the *Encyclopaedia Britannica*, yet costs nothing to its users and pays nothing to its writers. We see Google providing remarkably accurate answers to questions posed across an enormous web of information, based on nothing more than the right to run indexing software across that web and good technological decisions. What do we need to do to bring this functionality and efficiency to the sciences?

The US National Institutes of Health invests $28 billion annually in research. European spending is lower, though the EU is closing the gap, helped by a short-sighted US policy of flatlining scientific research budgets. Publicly funded basic scientific research often aims at eventual

commercial dissemination, but it is designed to fund the public good of research knowledge, the inputs on which all participants in a society, not just a specific company, can build.

The US investment in biomedical research is most often made available through the National Centre for Biotechnology Information, or the NCBI. At the NCBI website the human genome and associated information rest in the public domain, with powerful software and web interfaces available for users ranging from the expert molecular biologist to the citizen concerned about potential side effects to a drug. It is a remarkable system and it is truly free, both in cost to users and in intellectual property terms. However, the NCBI example is close to sui generis. Most disciplines do not have the NCBI nor the funding or expertise to create and sustain one.

Given that paying for research is not enough — that we have to get it to the scientists who might use it — we need to find more models beyond a large, powerful and centrally funded data operation. This is scientifically essential. In an interdisciplinary world, with unexpected connections the essence of scientific discovery, it is very hard to predict beforehand who the scientists are. That unpredictability argues for the minimum barriers to access, but it also implies more. A torrent of information can overwhelm as well as inform, and we need to structure access to the outputs of scientific research in a way that makes them useful in a way that allows the entire scientific community to build on them. We need, in short, to design the protocols of access to outputs so that we can build an open network of scientific discovery and commercialisation on top of them. We have a model for that network — it is the World Wide Web.

The outputs of scientific research come in a variety of forms. Again let us focus on biomedical research. Three of the most important outputs are: one, an article in a peer-reviewed scientific journal; two, a database that organises scientific data so as to permit multiple queries in future research; and three, research materials — cell lines, plasmids, reagents — assembled to test the hypotheses produced in the research. There are other outputs, clearly, but these three are hugely important. Each in its own way, if paid for by the taxpayer, represents public sector information. The research materials and tools represent hard-won knowledge encoded as usable tools, and they are too frequently left out of the discussion.

How are these outputs distributed? The open science recommendations from Science Commons argue that in each area we have paid inadequate attention to the legal, technical and social restrictions on distribution of outputs. We fund the inputs and then fail to adopt the best policies about making the outputs available. Most of the articles are behind paywalls — they cannot be integrated into a semantic web or linked and tagged for advanced search. The databases have contradictory legal and licensing requirements, incompatible structures and even different nomenclatures for the same gene or disease. Data-driven science works on integrating ever larger sets of databases; but we have built the data equivalent of the tower of Babel. Finally, the research materials are often inaccessible. We have neither a simple transactional system to get access to them nor an incentives system that gives reputational advantage to scientists who share rather than hoard their materials. This can be mitigated by good policy, which in turn allows good technology to be used.

National legal frameworks that understand the network and facilitate network-based innovation are an essential component of modern science, and Australia is poised to take a global leadership role in innovation policy. The federal Department of Innovation, Industry, Science and Research is reviewing the public sector's role in fostering innovation and developing a set of national innovation priorities, including identifying regulatory and other barriers to innovation and recommending ways to minimise them. This initiative could pave the way for a fundamental transformation in the way research is conducted. There is a significant and important opportunity to catapult the research in Australian science to impact levels disproportionately higher than in the past and increase the return on investment in scientific research. Australia is remarkably well positioned to implement these ideals. Not only is the national government interested, but there is interest at the state levels as well, as this hearing demonstrates.

Notably, there is unique capacity in Australia — essential, as the translation of these theories to practice requires skill and creativity. The Queensland University of Technology's OAK Law Project hosts the Creative Commons licences here and brings deep and lasting experience in the legal and policy work, as well as connections into the government information licensing framework and other projects looking at data licensing. In addition, the Australian federal research networks and university system have built a vibrant e-research community in the country, as evidenced at today's gathering here in Melbourne on that very topic. Australia is ready to lead in implementation, not just recommendation.

Creative Commons, a non-profit organisation based in the United States but with more than 45 international affiliates, provides tools and technologies that enable standardised, legally sound sharing and reuse of copyrighted materials on the Web. We have a strong presence in Australia, and our work represents a standard legal framework to manage copyrighted content into an open regime where rights of attribution can be reserved while guaranteeing to users a core right to use and build value on public sector information. As a non-profit, everything we do is free of charge and free to the world. In addition, we have a deep understanding of how to map intellectual property terms into the network, so that information available under our tools is searchable — by the rights granted to users — via Google, Yahoo, and many other major systems. We stand prepared to offer any and all assistance needed to state and federal governments interested in implementing the recommendations in the Cutler report on innovation.

Developing this critical new research infrastructure will leverage existing investments in building new research capabilities in Australia, as well as consolidating new models for the conduct of science and research which revolve around promoting the public good. Thus, we make four core recommendations: open access to publicly funded research and reports; open access to the data and databases collected with public funds; open access to the publicly funded research materials needed to build new experiments; and investment in cyber infrastructure for e-research.

The access to publicly funded research echoes the US mandates to scientists receiving funds from the National Institutes of Health, as well as many other such mandates from funders and universities worldwide. We also draw upon the NIH's policies on making research materials and related data available, though we go one step further in recommending a mandate on those two classes of public sector information.

We recommend these as default rules with a twist. The current system is one where the default rule is 'do not share', and it is exceedingly difficult to opt out. The individual scientist has a poor incentive to share, and potentially is punished for doing so as there are no rewards or even mechanisms to track sharing for eventual reward. Under our vision, the default rule is 'do share' but with simple and standard methods for opting out — to protect rights essential for commercialisation or for privacy and confidentiality reasons. This provides the best mix of settings so that only the information that truly deserves protection receives it, without encumbering basic information essentially by accident.

The good news is that Australia, which has already begun to explore using innovative legal tools for sharing information like meteorological data, has a unique opportunity to expand this exploration and take the lead in accelerating the pace of scientific discovery on a global scale.

This testimony sketches the outlines of principles — best practices — on access to public sector information broadly construed as text, such as government reports or government-funded scholarly articles; data, ranging from water quality to detailed scans of brain tissue and so on; and scientific research materials. Those principles share a common goal of getting us a more vigorous and innovative research commons. That is the public good for which our euros and dollars have already paid. At the very least this would be more efficient. At most it might produce in science the explosive growth in innovation and search that open networks have allowed in the digital domain.

The genius of the Web is that it is an open network. Anyone can link to any part of any page, and anyone else can link to that link. That web of citations and linkages is captured by search engines. We gain not only the knowledge in the content, but the knowledge supplied by those who read the content, who make connections the original author could not. It is this second layer of knowledge that assesses the first layer and makes searching it possible, which is something that scientists should understand. Peer review and citation play the same roles. But it is in this second layer that a world of fire-walled scientific knowledge will never develop, even if a line or two of the contents can be glimpsed from Google's search page.

The network is open in other ways. The architecture of the internet and of the Web is open and standard. That means that people can innovate without asking permission, can create new forms of communication and be sure they will run on the existing platform. Before the Web we had a series of closed networks, each controlled by a central authority — Ceefax, Minitel, CompuServe, the early AOL— each incompatible, unable to communicate well with each other. The explosive growth of the Web came about because it made the opposite design choices: open content and protocols, a focus on compatibility and interoperability.

There is an irony here. The World Wide Web was originally designed in a scientific lab to facilitate access to scientific knowledge. In every other area of life — commerce, social networking, buying books — it has been a smashing success. But in the world of science itself? With the virtues of an open Web all around us, we have proceeded to build an endless set of walled gardens, something that looks like Minitel rather than a World Wide Web for science.

The CHAIR — Thank you. Could I begin by taking up a couple of points from the paper you just presented. On page 2 you refer to QUT's OAK Law project bringing:

> deep and lasting experience in the legal and policy work, as well as connections into the Government Information Licensing Framework …

Then you made the point:

> In addition, the Australian federal research networks and university system have built a vibrant e-research community in the country, as evidenced at today's gathering here in Melbourne …

We were not privy to that; obviously we were here. I have never heard it painted in such glowing terms that we have a vibrant e-research community in this country. Given your international experience, what in particular is vibrant about our e-research community in Australia, and what are the points that you would recommend we look at particularly in our report?

Mr WILBANKS — It is rare for a country of this population to have the level of investment you have in things like the Australian National Data Service, and some of the things that are coming out of NCRIS and the other federal programs have really created a capacity to do big-scale computing, web-scale computing that is not very easy to do. The meeting of 400 or 500 people in the last couple of days on e-research in Australasia has been dominated by people from the Australian university system. You have projects on the remote control of sensor networks on the Great Barrier Reef; you have projects on large-scale data sharing and computational sharing that are the equal of what you would see in the EU and in the United States. I think it is punching well above its weight in terms of having a technical capacity.

Also we have many international affiliates at Creative Commons, and Australia is one of our strongest. What you have is a unique mixture of a strong and powerful capability technically and a strong capability legally. In many cases where we have a strong legal component there is no technical capacity in the country. If you take the United States, we have a very strong technical capacity certainly, but we have had some trouble on the IP side. Just getting the NIH mandate through was very difficult. It took years, and it is now under rearguard attack from publishers helping file new legislation and threatening to file lawsuits.

I think there is a really unique mixture of the two. It is really hard to understate how good a job is done up at OAK Law. It produces more good research reports than almost any of our affiliates, which is one of the reasons why I keep coming back to Australia. The more time I spend around the e-research community the more impressive it is. I have to come all the way around the world to find people who do the same work that we do at MIT.

The CHAIR — We have examined QUT's OAK Law project. Is there anything else you think is particularly impressive that we should make it our business to examine?

Mr WILBANKS — I think the Government Information Licensing Framework project, which is a Queensland project — predominantly Treasury.

The CHAIR — We are familiar with that.

Mr WILBANKS — That is an important project. In general the use of Creative Commons licensing is a policy decision that then has to be implemented. You have to figure out how you programmatically inject Creative Commons licences into documents or into datasets if you decide to adopt that.

The CHAIR — We took that evidence in Queensland. On the third page of your submission you state:

> We stand prepared to offer any and all assistance needed to state and federal governments interested in implementing the recommendations in the Cutler report.

Are there any particular recommendations in the Cutler report that you wish to highlight, and, based upon those, what do you think you and those around you could offer by way of assistance to implement those key recommendations?

Mr WILBANKS — The main one would be the use of Creative Commons licences on public sector information. There are choices to make in terms of which licence one uses. Creative Commons licences grant a basic right to the user to make copies. It is the primary grant you get from a CC licence. You can then add modifications that dictate whether or not commercial use can be made of the material, whether or not derivative work can be made of the material, and whether or not to impose what is called a share-alike or copy-left provision, in which you grant the right to make a derivative work but you impose the mandate to the user to put it back online under the same terms.

The CHAIR — So if the Victorian Government wanted to implement that, what in particular would you and your organisation be able to highlight to the Victorian Government?

Mr WILBANKS — We would make a point of sending people to provide testimony and guidance, and then there will be technical aspects, computational aspects, to implementing the licences, because they are available in three forms. One form is what we call lawyer-readable, and that is the binding legal code, but there is also a human-readable version of the licence for users who are not lawyers, and a machine-readable version of a licence that allows the file to be exposed.

The CHAIR — We did much work in that area in Queensland.

Mr WILBANKS — So we can help out with things like marking up documents, talking to the technical staff, making sure that you are able to use the licences once you have made the policy decision without having to spend a lot of time and effort on it.

The CHAIR — And would you have something over and above what is available and on offer from Queensland?

Mr WILBANKS — I think we would work through the Queensland group, but the core of our technical staff is in San Francisco, and the core law professors, other than the OAK Law folks, are based in the US, but we can make others available.

The CHAIR — So we could rest assured that we have the best on hand via Queensland?

Mr WILBANKS — Yes, and we would mainly work through OAK Law, yes.

Mr CRISP — Give us some examples of that. What institutions are using Science Commons at the moment?

Mr WILBANKS — Science Commons is just a piece of Creative Commons, so when I talk about Science Commons I am talking about Creative Commons primarily. Most of the licences are actually used at this point by publishers, so there are, as of this week, over a thousand peer-reviewed scientific journals that use Creative Commons licences in the world, and just a bit less than a third of all the open journals in the world use Creative Commons licences. To date the closest that a government has come to implementing Creative Commons licences globally would be the Australian recommendations and the work in Queensland. The Brazilian government has promulgated and promoted the use of the CC licences for cultural works that are paid for by public funds, but to my knowledge they have not implemented that by attaching the licences to every file that comes out.

Mr CRISP — Moving that forward, the adoption of Science Commons is going down the same path as the subset of Creative Commons that people are evaluating and rolling out and considering against open source and other forms of, I suppose, control?

Mr WILBANKS — Open source is primarily for software. Creative Commons licences are primarily for prose and for data in this context, so it is a complementary strategy that is more tailored to the content type.

The CHAIR — I have three questions, and I will pass them over, should you wish to just refresh your mind on what the questions are. The first is, what is the extent of support for Science Commons amongst universities here in Australia in particular and internationally, so we would appreciate both components being covered. Secondly, will the adoption of Science Commons have an impact on the ability of universities to extract commercial value from their research? And that is presuming you think that is not a bad idea; you may wish to make a comment on that. And in Canberra our committee heard from Professor Richard Jefferson, head of the BiOS initiative where patented information can be shared on neutral terms. How does the Science Commons differ from or overlap with that system?

Mr WILBANKS — The first question is the extent of support for Science Commons amongst universities here. The primary support we have now is a reflection of the special relationship that OAK Law and Creative Commons Australia have with QUT. So we have really begun our discussions with and at QUT, and I was a special guest of the VC all this year, and that is actually what paid for my trip here today.

The CHAIR — The VC?

Mr WILBANKS — The Vice-Chancellor, Peter Coaldrake. We have worked most closely with them. We have also over the last six months begun to develop a very good relationship as well with the University of Queensland, primarily their e-research and library groups. One of the core research groups I was talking about in my testimony is the e-research group at the University of Queensland, which has a remarkable computing effort that really parallels what we do at Science Commons, so what we are talking about now is a three-way partnership between Science Commons, QUT and UQ that would leverage the policy work of OAK Law with some of the science and technology work happening in the Institute for Molecular Biology and Research at UQ, and that is the core of the work that we have done to date. This trip

is the first time that we have really branched out and started talking to the universities in Sydney and Melbourne. In the coming months we will have deeper relationships built, starting with Monash especially, which is the host of the Australian National Data Service. Secondly, will the adoption of Science Commons have an impact on the ability of universities to extract commercial value? We hope it will increase the overall amount of value created and not impact the ability to extract revenues, so we view what we are doing as fundamentally compatible with commercialisation. The vast majority of — —

The CHAIR — Can you expand on that and why you would make such a comment, and while you are doing that I am going to check and see if I kept something I tore out of the paper this morning that relates directly to that question.

Mr WILBANKS — Sure. There is a frequent assumption that open science is not compatible with commercialisation or with patenting and that you really have to go out of your way to prevent that sort of exploitation of public funds, taking of public goods — the rhetoric can be quite overheated — but the reality is that the markets are quite good at doing some things, and turning some of these basic research results into valuable products for people is something that markets are good at and we have to preserve the ability of the markets to do that.

The question really is the methodology by which we approach that, and the way we do that right now is we try to lock up all of the small pieces, which then creates very high transaction costs and makes it fairly hard to be an entrepreneur, especially in the life sciences because you really do not know what you are doing in the life sciences. Despite all the rhetoric, it is more like trying to do a moon shot when you do not understand physics. That leads to this instinct to sort of patent all the little stuff. What we are hoping to do is to first allow for much lower cost of asking risky questions. So if you have access to all the data that has come before, all the materials that have come before, you can very rapidly and at a low cost test wild theories, whereas right now it is quite expensive to do that and you get punished for it, because if you do not generate papers quickly and you do not generate data quickly, you do not get your next grant, and if you are an entrepreneur and you have got venture capital funding or state funding, you are biased into asking sort of small questions incrementally, getting a patent and then getting revenues out of that. If we can actually lower the cost of collaboration and increase the number of collaborations and the speed at which they happen, it is one of the only mathematically sound ways we have to increase the chances of big discoveries coming up.

This is why I talked about a default rule of do share where it is easy to opt out, because it means the minute you identify something of value, you can say, 'Okay, now this goes back behind the firewall, because we need to have the commercialisation rights to it'. But the way we do the system right now is everything is assumed to be valuable and kept behind the firewall and you have to make an active decision to move it out into the public, and the argument is that if we can change that and you have to make an active decision to take it out of the public and you make it easy and standard to do that, that is the right balance of settings, because that preserves the right to commercialise whenever you think you have the need to commercialise without accidentally locking up the vast majority of things that do not need to be commercialised. There is also a real benefit to the companies of not having to negotiate the agreements over and over again, so if you make, for example, a wide pool of these materials available under a contract that has already been approved for research purposes, all the company has to do is say, 'Oh, those are available under a standard licence, sure. We do not have to negotiate that; we do not have to pay our lawyers; you do not have to pay your lawyers'. And you basically delay and you only negotiate and spend the money negotiating when you know you need to, so it is actually a much more efficient use of everybody's time, but that should be fundamentally compatible with commercialisation.

In answer to the last question, about Richard and the BiOS initiative, Richard focuses on what I would call the capacity to practice or the capacity to take action and take something out into the market and the incumbrance that the patent mosaic provides on that. I think he has done some of the best work out there on really demonstrating that there is a thicket of patents on almost any

important biological activity. I think in particular his work on making the patents transparent by making them searchable and understandable is really important work.

We do not really work in the patent space actually. We work primarily under the idea that there is a layer before you commercialise, which is the research layer, and that activity at that layer is tremendously inefficient; basically it is like an engine that is running at 10 per cent speed. If we are successful and we can get the engine running at a much higher speed, or at a much higher percentage of its capabilities, it is going to kick out, ironically, probably more patents, not less, and that will increase both the pressure on people filing those patents to understand the landscape and on the owners of those patents to think about how they are going to make use of them. My hope would be that as we make the research layer more efficient and unis start getting more of those patents they begin to exercise the power they have in those patents in a way that facilitates both research and social good as well as commercialisation — to make them available under something like Richard's concordance, or a licence that grants expansive research rights to the world, or commercial rights for people working on developing nations' issues of hunger or malnutrition, water quality or rare diseases. Because you can grant those rights while retaining the right to make a lot of money as well.

The CHAIR — Have you got examples of that?

Mr WILBANKS — The University of California is probably my favourite example. Rather than adopting a standard licence or a constraining single-form contract, they have a set of principles for what is called socially responsible licensing, and those principles involve granting research rights for their technologies, because if you have got a patent having other people do research does not degrade the value of the patent at all — because you have got other people improving your property for you, right? Secondly, those principles recognise the university has a role to play in the social value of the technologies that are created there.

So if they come up with a drug at the University of California, the goal is not to licence it in a way that maximises cash but in a way that maximises the balance of cash and social benefit. They would proactively grant rights to have manufacturers make generic versions of that drug for leishmaniasis but they would reserve the right to sell it to travellers who wanted to make sure that they did not get sick with giardia. So you can do that by keeping principles that you encode in your contracting process that are relatively easy to do, and that really relieves a lot of the burden on the patent system for those who cannot afford to negotiate through. If you have a cure for Alzheimer's disease, your lawyers will figure out how to negotiate the rights you need. If you have a cure for the cold, you will figure out how to negotiate it. But if you are trying to come up with a cure for sleeping sickness, it is not going to be possible. So by a socially responsible practice you can do that.

There is a licence called the equitable access licence, EAL, which attempts to actually impose downstream behaviour on anybody who licences the technology. We have not gotten into that discussion because we tend to not believe in using contracts to reach through and impose behaviour on people. That is why we like Richard's concordance more than that. There is also something called the open source drug discovery effort which is coming out of India, which has a click-wrap patent licence again. It is like a patent-left or a GPL for patents. I think Richard has probably given you the testimony on why that is really hard, which is that someone who does not sign it still has the right to sue people who did sign it. There can be patents that are held all over the world that impact your patent, so your licensing of that patent is frequently not enough. That is why we buy into Richard's transparency and concordance there.

The CHAIR — I wanted to just develop that further. The paper I was looking for this morning was the *Age*, which had a section on the Australian Stem Cell Centre. I am pretty sure it was this morning's paper; if it was not this morning's, it was yesterday's. It talked about a centre of excellence in Australia that looks at stem cell research. According to the scientific community engaged in stem cell research, it has received appropriate funding. According to people in any

other field of research, it is currently the flavour of the month and it has received a disproportionate amount of funding. So there is that rivalry within the research community. Then there is the rivalry between the adult stem cell sector and embryonic stem cell research, and there is another layer, of course, of destructive embryonic stem cell research.

The centre is based on a collaborative framework with a number of Australian universities. To cut a very long story short, the CEO has gone, the board has been replaced, and it is in an interesting position. While you were talking I was wondering whether the Creative Commons, the Science Commons, would have been any better for the future of stem cell research and collaboration. You talked about only 10 per cent of the research really in most institutions being effective because it is all locked up. Give me an example, other than the University of California, of the kinds of things that work really well. And do they cut down rivalry between researchers? Do they assist in collaboration between research institutions?

Mr WILBANKS — Sometimes.

The CHAIR — Because you paint a very rosy picture.

Mr WILBANKS — Of course; that is my job.

The CHAIR — What would work in the scenario that I have just outlined that, had it been more in the Science Commons mould, perhaps would not have occurred?

Mr WILBANKS — The law and technology are the easy part. It is easy to design contracts that create open systems; it is easy to design technologies. It is very hard to overcome social and institutional traditions. That is the hardest part about all of this. That is the hardest part about implementation.

The CHAIR — This site has been going for under a decade.

Mr WILBANKS — Yes.

The CHAIR — So it has not got a long tradition.

Mr WILBANKS — It has a long tradition in its own way. The classic crack of Newton's was: 'If I have seen farther than others it is because I stand on the shoulders of giants'. That is taken as a way to justify sharing in science but actually it was an insult to his competitor, who was a short man — Hooke. So scientific competition, rivalry and insult is part of science; it is not going to be taken out. What we can do is look at a couple of examples of where the culture was changed by a sort of collateral force. One would be the human genome.

When the human genome project began genome research was done by individual labs and rooms like this, and it was one gene, one lab, and the first one to publish sort of won the lottery. The federal government in the US mandated public domain status but no-one was depositing. Then a private company came along, Celera, that competed, which scared Congress into getting its teeth into the mandate. The publication community required in any publication about a gene that the gene have a database ID number, and you could not get the database ID number without putting the whole sequence into the database in the public domain.

This completely changed the structure of genetics and genetic academic publishing permanently, but it took the combined influence of the publishers, because the publishers provide the money basically to the scientists. You do not get cash as an academic scientist; you collect citations, and those get you grants and jobs. The publishers getting involved created a powerful scientific incentive to play ball, and the government getting involved created an economic incentive, which is, 'You are not going to get more funding'. One lesson to be drawn is to engage the people who control the currency of the realm, which is publications and grants. That is where the real leverage and the power is.

As to the stem cell, I do not know if our proposals would have saved something that is so political at heart. If the argument is over whether or not the funding was too big, arguments over how to manage the transaction cost and technology probably are less important. What I do think we can start to think about and provide, though, is: how do you provide rewards to those who share? How do you provide an incentive for sharing? One way you can do that is by thinking of how you get rewarded in science already for sharing, and that is, again, through citation — that is, when you publish, you share to a certain extent. If you were able to ask, 'How many times was my dataset accessed? How many times was my dataset cited? How many times was my material promulgated? How many times was it cited? How many papers came out of that data and those materials', and actually track that and provide those as a rating factor for scientists in addition to citations, then you could actually start to create some pretty powerful social incentives for scientists to comply with. It has to start with the creation of multidimensional ways to track the impact of public sector information. You can do that through hyperlinks that are going into a dataset, you can do that through the number of times it is downloaded, you can do that by the number of times it has been cited, but it has to start from the bottom and come up by the creation of these things. That is the second part of the question.

The third part of the question is: are there places where this is working? The places that we study for this — and these are the places where the bulk of our funding comes from — are in the US. They are called disease foundations. This is a group of individuals, typically very high net worth individuals or coalitions of lots and lots of individuals, who collectively raise money to fund bespoke research into a given disease. We have in particular focused on brain diseases, because there are a lot of brain diseases that do not have enough people to justify modern pharmaceutical research.

Multiple sclerosis is a good example for a lot of reasons. There is foundation now called the Myelin Repair Foundation that is trying to answer five questions about multiple sclerosis that are considered essential to understanding the disease. They have provided an enormous amount of funding to a small set of universities to do this research and answer these questions, but they are mandating pre-publication information sharing among anyone who takes the money. It took them a long time of negotiating with the tech transfer officers at those universities. Do you know these folks?

The CHAIR — No; I like the idea.

Mr WILBANKS — It took them a long time, and it is something like a 150-page collaborative research agreement that each university signs with the foundation, and it has created this little zone of sharing around this little zone of content, and to all accounts it is working extremely well, and it is run by former pharmaceutical executives.

There are four or five of these examples I can give you. There is one in Huntington's disease, called the Cure Huntington's Disease Initiative — CHDI. They are my primary funder. There are a few others that are doing the same sort of thing. They run into constant traffic from the universities who do not want to accept these sharing agreements, and they cannot connect to each other. I referred to walled gardens in my testimony. If you think about AOL and CompuServe and Prodigy in 1991, if you used AOL and I used CompuServe we could barely talk to each other. Despite the fact that all these folks are doing research on the brain and they all want to share, the myelin repair network does not connect to the Huntington's network, and they are looking at the same cells, the same genes and the same proteins.

A huge part of what we are trying to do is to take what they are doing, but do it in an interoperable way, because we are starting to see real results in these enforced sharing gardens that are taking off. The Gates Foundation is doing this now for HIV and malaria research. They say, 'We will not give you money if you do not share'.

A huge part of what we do at Science Commons is try to provide standardised ways to implement so that anyone who actually wants to buy into this is connected to anyone else who buys into it, whereas right now everyone writes their own network. It is both very encouraging that it is working and very frustrating that we are basically replicating failed closed-network models of the past. You get these wonderful benefits inside the network, but I have to basically be a university and negotiate my way in to see your stuff. If two people are working on the same problem in different places and they do not go to the same conferences or read the same papers, they will never know, so it is a limited efficiency game.

The CHAIR — Thank you. That is very helpful. Is there anything you want to cover?

Mr CRISP — I have covered everything I was thinking about and more.

The CHAIR — Okay. The one thing we have not perhaps covered is the resistance by the scientific and research communities. I think you have probably covered that enough, unless others want to delve further.

Mr WILBANKS — I can provide Vaughn with some links to some of the empirical research on resistance and how to overcome it.

The CHAIR — Okay. Thank you very much. That has been really interesting and very helpful. Mr Wilbanks, we will be providing copies of the transcript to you in about a fortnight, in whichever part of the world you are.

Mr WILBANKS — I will be home.

The CHAIR — That makes it easy to track you down! You are free to correct typographical errors but not errors of substance.

Mr WILBANKS — I do not think I said anything too bad today!

The CHAIR — I am sure you would not have made a single mistake, so thank you very much.

Mr WILBANKS — Thank you.

Witness withdrew.