



COMMONWEALTH OF AUSTRALIA

Official Committee Hansard

JOINT COMMITTEE ON PUBLICATIONS

(Briefing)

Reference: Certain issues related to non-print material

THURSDAY, 17 FEBRUARY 2000

CANBERRA

BY AUTHORITY OF THE PARLIAMENT

INTERNET

The Proof and Official Hansard transcripts of Senate committee hearings, some House of Representatives committee hearings and some joint committee hearings are available on the Internet. Some House of Representatives committees and some joint committees make available only Official Hansard transcripts.

The Internet address is <http://www.aph.gov.au/hansard>

To search the parliamentary database, go to: <http://search.aph.gov.au>

JOINT COMMITTEE ON PUBLICATIONS

Thursday, 17 February 2000

Members: Mr Lieberman (*Chair*), Senators Bishop, Calvert, Chapman, Lightfoot, Ludwig, McKiernan and McLucas and Mr Hardgrave, Mrs Hull, Mr Lloyd, Ms Jann McFarlane, Mr Rudd and Mr Sidebottom

Senators and members in attendance: Senator McLucas and Mr Hardgrave, Mr Lieberman, Mr Lloyd and Ms Jann McFarlane

Terms of reference for the inquiry:

Briefing on certain issues related to non-print material.

WITNESSES

NICHOLS, Mr George Ernest, Director-General, National Archives of Australia9

STUCKEY, Mr Stephen John, Assistant Director-General, Government Services, National Archives of Australia.....9

Committee met at 8.03 a.m.**NICHOLS, Mr George Ernest, Director-General, National Archives of Australia****STUCKEY, Mr Stephen John, Assistant Director-General, Government Services, National Archives of Australia**

CHAIR—Good morning, ladies and gentlemen, and welcome to the first meeting of the Joint Committee on Publications for 2000. Today we are going to start with the private briefing and then quickly at the end we will deal with our formal matters so we can report to the Senate and House of Representatives today. The committee is undertaking a series of private briefings prior to launching a full inquiry into certain issues relating to the kinds of non-print material authored by government and parliamentary sources. In particular, the committee is seeking information on access to such material by Australians in remote and regional Australia and on preservation of such material. The committee will not be restricting itself to material published on the Internet. Instead, it intends to investigate a variety of electronic formats, including CD-ROM and newer technologies such as digital and video disc, as well as material recorded on older technologies such as audio cassette and videotape.

Today I welcome officers from the National Archives of Australia, which is the second briefing for the committee on this particular subject. While we are going to be very informal there are a couple of formal things I have to put on record that are always put on record when parliamentary committees are receiving statements and submissions from members of public and officers. I remind you that proceedings here today are legal proceedings of the parliament and warrant the same respect as proceedings in the Houses themselves. The deliberate misleading of the committee may be regarded as a contempt of parliament. I would now ask George and Stephen to make an opening statement before we proceed to questions.

Mr Nichols—Thanks, Mr Chairman. To assist the committee, I will just make a few brief remarks and then ask Steve to talk a bit in more detail. We have tried to tailor our remarks to what we think is directly of interest to you. I will make a few introductory remarks just about the archives. Our actual collecting mandate relates to government records primarily. There is not very much outside that. We do not generally involve ourselves in printed published material. However, with the advent of electronic media of various sorts, the distinction between printed material and original records is a bit blurred regarding when something is published and when it is not. I do not want to dwell on that because I do not think in the end it makes a lot of difference, but it does for us in terms of our collecting mandate, say, vis-a-vis the National Library. We draw try to draw a line there.

In terms of electronic media, the issues, whether it be printed, published material or indeed original records, tend to be the same. It is a subject that has concerned us for a number of years as you would expect. It is a subject where there are no answers. We have had to try to develop our own answers. Our contacts with national archives overseas have yielded the same thing; they are no more advanced than we are. In fact, we think in many cases they are behind us. We do keep in contact with the major national archives overseas and share information on this subject. What we will be talking about is best practice, as far as we know, in archives around the world.

Briefly, there are strengths and weaknesses in these formats. The issues that concern us are about capture of information, preservation of information and accessibility over the long term. We see all those things as problems. On the other hand, the great strengths of some of these electronic media are storage, copying, access in the short term and portability. So you get advantages and you get disadvantages – it is just a conundrum. The weaknesses are things like fragility, unreliability of the medium you are using and the difficulties that you encounter because of the need for migration of material. The thing that very much concerns us is that we are here to provide long-term accessibility; we are talking about giving access to things in 100 years time. It is very difficult to devise ways and see ahead as to how you can ensure that, as opposed to with a much more physical format like paper where we know where we stand.

That has forced us to fundamentally reappraise our own approach to the business we are in of archives and archiving and to go back to first principles. A general statement I would make is that the new approaches and the principles and standards we have developed are not media specific. Format, whether it is paper or digital format, does not matter to us in that sense. The principles we now apply to trying to handle records are irrespective of format. I thought it might be useful for Steve to talk about some of the products we have developed and standards we have reached.

Mr Stuckey—We in Australia in the archives and records community and in the software industry as well have embraced standards for records and publication. Australia is the only country in the world that has a standard on records management. That standard gives a framework in management of all records irrespective of format, and that has been brought about by the digital environment in which we live. The digital environment allows you to keep everything. You can keep everything that humanity creates and you can keep it in a relatively small space. Increasingly, compression is allowing you to keep more and more in the same space. So storage is not an issue.

Finding information and linking information so that you have a logical progression is the issue, and it requires major front end input. If you do not plan for accessibility and retention at the beginning, it just will not happen. We find that with a lot of early computer data. A lot of the data for NASA, for instance, is just not accessible any more, nor are some of the census records in the states. In Australia, a lot of the petroleum data created in the 1960s is no longer available because it was on old analog or very early digital computer tape.

10

What we have done in the Australian information industry – the information industry is all about libraries, archives and the software industry – is to develop standards for metadata, which is information about information. Metadata is the sort of stuff that you implant in a record as you create it, such as who you are, what it is about, when it was created, who has access and what it is linked to. That is to ensure that you can find the data in the future and that you can link like data. It is no good my sending an electronic document to somebody and storing that as one entity and the reply coming back and being stored as a totally separate entity, which is what happens unless you link the two. That is what metadata is about.

Migration is also an issue – migrating data forward to new standards of software and hardware. Given that software these days has a life expectancy of 24 to 30 months, there is in effect a rewrite of software every two to 2½ years and forward compatibility is not always guaranteed. So we are putting in place standards with the software industry to ensure that forward compatibility is always possible.

Development of proprietary software is always an issue – proprietary software being where somebody sets out to develop their own software and does not buy something off the shelf. We need to know where that happens as far as possible. Every time you convert a record from one software type to another software type you lose a certain amount of the functionality. It is extraordinarily expensive to ensure that you move 100 per cent of data from one version to the next version. So we are faced with the question of what is an acceptable loss in functionality and when does a publication no longer become a reliable, authentic, evidentiary publication or record. So that is an issue to us.

Conversion from analog format, whether it be analog tape or this sort of stuff – paper – to digital is extraordinarily expensive. The ABC is finding that converting its analog video to digital costs about \$400 an hour of broadcast time. Conversion from sound to CD is much cheaper – about \$90 an hour – but it is an expense and therefore people have to decide what they are going to convert. But once you have converted then copies are extraordinarily cheap. It costs less than \$1 to duplicate a CD. Once you have converted, you can make stuff available cheaply and very widely.

Long-term access is also an issue with certain types of records. For example, how do you mask part of an electronic record for privacy concerns? On paper it is easy: you can copy it and white it out. But it is no good putting the whiteout on the screen or whatever. So they are issues to us.

What the archives has done is to establish in consultation with other government agencies – for instance, AusInfo, the Office for Government Online and the National Library – and the software industry a range of standards for records creation and records accessibility. This government and the previous government pushed very hard for Australia to undertake business and government online. So we have been pushing very hard and have gone a long way to ensure that publications and information are available online to all Australians. But how do they find what they want? Consequently, we have written a standard about metadata. We have written a range of thesauruses so that, if I am talking about kindergartens and you are talking about early child care and early child education, we both end up at the same place. We have just issued a thesaurus of 300,000 common terms and converted those to bureaucratese, which is the way that government papers are often written. This will enable people to go to information kiosks, for instance, and access information about government and services to government.

We have also provided advice to government agencies about how they manage their electronic records and their electronic publications, whether it is email or it is just an electronic record. What we have in final draft form now is a guideline and a standard for government agencies on how they archive their web sites. This has been developed in consultation with AusInfo and the National Library. This is important for a range of reasons: a web site contains important publications, but these days people go to web sites to find out about what services government can provide to them. Therefore, the agencies have a duty of care to ensure that that information is correct. If you feel that you have been denied an entitlement, for instance, then some years later you could go back and say, 'The information that was provided to me on the web site led me to believe that I was not entitled to that.' So agencies have to be able to bring back what that web site looked like on 22 June 1996 for evidentiary purposes.

As George said, we do not care whether it is a publication or whether it is a record now. Certainly, the people who use this stuff do not care. Consequently, we have developed our standards with the National Library and with the Australian government publishing arm – AusInfo. I think that is about all we need to say.

CHAIR—It is fascinating. We are going to have time for a few questions. Then there will be some questions that we will not put to you, but the committee would like to write to you so that you can respond on those as well.

Senator McLUCAS—Nothing jumps to mind, thank you, Chair. But, yes, I agree that it is a great challenge.

Mr HARDGRAVE—One thing that came to mind was the question of authorship. Are you confident when you take a record at face value that you are right? What happens when somebody says, 'Here you go, here is all of my intellectual property,' and it is not?

Mr Stuckey—You have to take some things on trust but you also can design a system, if it is a network, so that, if I log on to the system, the system knows who I am, whether I know they know who I am or not. It is difficult, however, say for a poet or an author who could go to the National Library and say, 'Here is my work on a disk.' You have to take that sort of thing on trust.

CHAIR—Produce your drivers licence maybe. It is an interesting issue.

Ms JANN McFARLANE—Do you have all our web site and parliamentary web site information on your system?

Mr Stuckey—We do not, but the parliament does. I think the Parliamentary Library and/or PISO do, but I am not sure that they have kept every version.

Ms JANN McFARLANE—That was one of the things we explored last time, wasn't it?

CHAIR—That is right.

Ms JANN McFARLANE—So your link to us is via the Parliamentary Library. In terms of national archiving, the Parliamentary Library is responsible for retaining all our web site information.

Mr Stuckey—The library is responsible for keeping its own information. The records of the parliament are Commonwealth records, but the parliament under regulation of both Houses looks after most of its own information.

Mr Nichols—That is into the long term as well. We do store some paper based formats for both houses of parliament but, in the main, it is kept here.

CHAIR—Could I flag on that point that we will write to you and ask if you could give us a list of what you store for the parliament. Is there a protocol and an agreement that identifies those things or has it just grown over a period of time? We would be interested to know that.

Mr Stuckey—Yes, they are covered by regulations.

Ms JANN McFARLANE—I was not sure of the links between what the National Archives kept and what was our responsibility. You said that you talk to other bodies like AusInfo, which comes under our umbrella, on the preservation side or on the developing of protocols for this metadata for making sure that, when it is kept, it is being able to be transformed and retrieved some years later.

Mr Stuckey—Yes. AusInfo was one of the agencies that worked with us on establishing that metadata.

CHAIR—Do you make any special provision for people with disabilities in the way that they might seek to access archival records?

Mr Nichols—Are you talking about electronic records or paper records?

CHAIR—Probably both. You might like to take it on notice.

Mr Nichols—In terms of paper records, our reading rooms are accessible. The problem you encounter is with more remote access. In other words, it is fine if you live in a capital city but, if you do not, you have to travel to get to a capital city to access material. That is where the Internet is becoming a much more powerful tool. It is our primary point of reference for people coming to the Archives now.

CHAIR—It is opening up an enormous new opportunity for people with disabilities.

Mr Nichols—Yes, or it can just be remoteness rather than any physical disability.

CHAIR—Sure, physical and other.

Mr Nichols—Our primary database is now available on the Internet so that you can do your initial searching there before you ever come near us. Obviously, the possibilities open up further that you could request a document which we could image and transmit to your living room. We do not do that at the moment, but technically you could.

CHAIR—This has been a very brief session but I have been absolutely fascinated by the outline you have given us and, as I have foreshadowed, the committee will be writing with some other questions that can be more appropriately answered by letter. I would like to congratulate you for leading the world on some of these technological changes. I think that is great. One concern we would have with what is being done so far as

interfacing with everybody else is that duplication does not occur. If you are refining protocols that can be used by other agencies, we would be particularly interested to see whether we are and, if we are not, how we might improve that. I congratulate you and wish you well in your work.

