



COMMONWEALTH OF AUSTRALIA

# Official Committee Hansard

JOINT STANDING COMMITTEE ON PUBLICATIONS

(Briefing)

**Reference: Certain issues related to non-print material**

THURSDAY, 9 DECEMBER 1999

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 OF PUBLIC ACCOUNTS AND AUDIT**

**Thursday, 9 December 1999**

**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 Ludwig and Ms Jann McFarlane, Mr Hardgrave, Mr Lieberman and Mr Sidebottom

**Terms of reference for the inquiry:**

Briefing on certain issues related to non-print material.

**WITNESSES**

**CAMERON, Mrs Jasmine, Manager, PANDORA Project, National Library of Australia..... 1**

**CATHRO, Dr Warwick Scott, Assistant Director-General, Information Technology, National Library of Australia..... 1**

**FULLERTON, Ms Jan, Director-General, National Library of Australia..... 1**

**WEBB, Mr Colin, Director, Preservation Services Branch, National Library of Australia ..... 1**



**Committee met at 8.00 a.m.**

**CAMERON, Mrs Jasmine, Manager, PANDORA Project, National Library of Australia**

**CATHRO, Dr Warwick Scott, Assistant Director-General, Information Technology, National Library of Australia**

**FULLERTON, Ms Jan, Director-General, National Library of Australia**

**WEBB, Mr Colin, Director, Preservation Services Branch, National Library of Australia**

**CHAIR**—Welcome. This is a private briefing on the issue of non-print material. It is not an issue that would normally get the adrenaline running—

**Ms Fullerton**—It does for us.

**CHAIR**—And it does for us too, because it is a fascinating issue. The committee is undertaking a series of private briefings prior to launching an inquiry into certain issues related 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. I might also add that we are particularly interested in how people with disabilities are being served and assisted or, perhaps, disadvantaged, and how we might overcome any disadvantage in that area.

We are also seeking information on the 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, newer technologies such as digital video disc, as well as material recorded on older technologies such as audio cassette and videotape.

This morning I welcome officers of the National Library of Australia to kick off, significantly, our first round of briefings. So you can see the status and importance we attach to your advice. Members of the committee are here at the moment, as you can see, and I welcome their presence and assistance at the beginning of what I hope will be a very worthwhile inquiry and one which will be of benefit to all people involved in this fascinating issue.

There is one formal matter that all committees of the parliament are required to place on record, and I am sure you have heard it all before. I remind everyone that proceedings here today are legal proceedings of the parliament and warrant the same respect as proceedings in the houses themselves, and any deliberate misleading of the committee may be regarded as a contempt of parliament.

Ms Fullerton would you like to make an opening statement before we proceed to questions, bearing in mind that we will try to end at about 8.20 a.m. At the end of that time the committee has some formal matters to attend to which it must do because the parliament expects us to report to them today on some matters.

**Ms Fullerton**—I will make a very brief introductory statement. The National Library does have a profound interest in this topic because we are charged with ensuring that our

---

documentary heritage is preserved and accessible for posterity. We have been trying to grapple with this issue of our heritage, our culture, our information in electronic form, for a number of years. It is a wonderful distribution and access method, but it presents real problems of preservation and ensuring that things are accessible forever.

There are a whole series of other problems to do with the authentication of the material, making sure that the electronic document is the same document, is unchanged, that there is control of versions, editions, and that you do not have one version up on the web and when you want to produce a new one you just take that one away and throw it away, which is a very common thing that is happening in this era of the web. The other forms of electronic technology present problems that are almost equally as great in magnitude. Perhaps my colleague, Dr Warwick Cathro, can talk about some of the things the National Library is attempting to do.

**Dr Cathro**—I have a brief handout here. Perhaps I can table it. It is only two sides of one page. I will speak to it anyway.

**CHAIR**—We will receive that. At this stage it is a private briefing. Any documents tendered will be treated as background material and there is no resolution required by members today.

**Dr Cathro**—Given the constraints of time, I will just speak to some of this. This is to inform the committee about the digital services project which the National Library is currently undertaking. Part of the background to this is that the National Library, under its act, is required to collect and preserve information in all formats and to cooperate with others. It also has to deal with the digital issues in the world of digital publications. There has been a big transformation in the 1990s. A significant proportion of Australia's cultural business and other information is now made available through the web, and in many cases only through the World Wide Web. The Library's directions statement takes this into account, as Jan has mentioned.

There are two types of digital resource that often have different requirements. The first type is those that are created originally, which are born digitally, if you like. An example of those is the web sites that the National Library is currently archiving on a selective basis through its PANDORA Project. Then there are those that have been digitised or converted from other formats, such as the pictorial collections in the National Library that we have digitised. We are also proceeding to look at other areas—rare maps, key manuscript items and so on. We want, in a sense, to collect these materials into a digital library so they can be managed.

The same kinds of challenges or issues we have with traditional libraries in selecting, acquiring, cataloguing and preserving the information arise with digital libraries. In the background material I have mentioned a number of challenges that libraries have to deal with, and a key one is the preservation challenge.

Colin Webb may have an opportunity to enlarge on this but, because of the rapid changes in technology, information in current formats is going to become obsolete relatively quickly. Unlike paper, even acid based paper, which can survive in libraries for hundreds of years or whatever, there is a much bigger challenge with digital information because of the rapid obsolescence and therefore the problem of how we are going to allow Australians in 100 years time to read the information which is produced in digital form today.

**CHAIR**—Warwick, I have often wondered whether the ancient parchments last longer, in fact, than the acid paper. We will not know, I suppose, for several thousand years.

**Mr Webb**—They do.

**CHAIR**—It is a curiosity of the chairman's.

**Dr Cathro**—Because of the range of challenges, including the preservation challenge, that the National Library has, in order to fulfil its own responsibilities it has embarked on this digital services project, which is an IT project, if you like, but one that will allow the Library to grapple with the challenges in its strategies.

We are currently at a stage where we have chosen a product to support the cataloguing and searching of digital information and to fit with the other products we have such as our catalogue and our Kinetica service, which is a national service used by libraries. We are currently evaluating tenders for a digital collection management system which will allow the digital collections I have mentioned to be managed and developed, and certainly our aim would be to allow them to be migrated to new formats as technology changes.

Currently, we are still at the evaluation stage, but we see the project as a key strategy to achieve the Library's objectives in the area of digital information. While it will not be a magic bullet to solve all of our problems, it is going to give us a better capacity to address these difficult challenges. Our goal is to ensure that the digital publications of today are preserved for citizens of the future.

**CHAIR**—Thank you, Warwick.

**Mrs Cameron**—I will briefly tell you something about the PANDORA Project. In 1996, when the Internet was fairly commonplace in the university sector but only really emerging into the general community, we realised that the academic sector was actually publishing on the Internet and, as Jan said, because of our preservation role, we felt we had to move to try and do something about it. It really was a learning curve for the Library for the first year or so. We looked around at the other national libraries, and there was very little being done. In fact, the Library of Congress still has not addressed the preservation of digital publications. We found that Canada, Sweden and the Netherlands were really the only three other national libraries that had started to look at this issue. We developed a set of business principles. One of the key business principles that we still work by is that, if the Australian title on the Internet exists in print, we do not capture it for PANDORA. We will only capture publications that are what we call 'born digital' in online form only. That really does cut out a large amount, particularly of government publishing. What it does mean is that in PANDORA we capture at the moment academic titles. There is quite a bit of original online publishing going on in the university sector, and there are a lot of community groups and what we call self-publishers using the Internet—people who want to put out a particular message on a topic. We try and cover topical issues like the gun debate, the Pauline Hanson debate, the euthanasia debate. We will specifically search the Internet for topical issues.

**CHAIR**—Do you include the reports of this auspicious committee?

**Mrs Cameron**—We will if it is online only.

**CHAIR**—We are relieved to hear that.

**Mrs Cameron**—I have to say that, because of the complexities of preservation, we still go for the print in every instance if it is available. What we have noticed with both the federal and state governments is that print runs are becoming smaller as government departments are putting up alternative access on the Internet, so we have to be really quick if we want to get a print copy of many government publications. You get print runs as small as 200 or 250, instead of in the thousands. If we miss out, we will take that title in online form.

One of the issues for us as well is that online material is not covered by the legal deposit section of the Copyright Act which, as you know, is currently under revision, and the Library is contributing to that debate. It means that every time we search the Internet and find something that we think—and it is quite a subjective judgment—is worth while to archive, we negotiate with the publisher. We have to seek permission; if we do not receive permission we do not go ahead and take that title into the archives. It is a matter of negotiating and seeking permission.

**CHAIR**—Is that because you think that is a legal requirement? Is that your intuition, that the law might—

**Mrs Cameron**—It is a legal requirement. We cannot just capture their information.

**Ms Fullerton**—Without their approval.

**CHAIR**—Okay.

**Ms JANN McFARLANE**—Has anyone ever said no?

**Mrs Cameron**—Very few. We have had a couple of noes. The two instances that I can remember were from Aboriginal groups that had put material up on the Internet. One group was rather suspicious of what we were trying to do and would not really discuss it with us. The other group just felt their material was culturally sensitive. They will not give us permission at this stage but they are still thinking about it—we are negotiating with them. This is one of the many issues we face with capturing material on the Internet.

We really need a system, which we do not have but we are seeking under the digital services project, whereby we can restrict access to certain types of material, whether they are culturally sensitive or fall into other categories. There is a distinct difference. What we are doing at the moment is using a software robot to go out and capture publications on the Internet and bring them back, and we store them on the Library server. That is really the archiving function. But there is a whole preservation side that is yet quite undeveloped. The Library is probably almost leading the world, with Sweden and the Netherlands, in looking at the preservation techniques you can use for migrating or moving these publications forward as the software they use changes, things like that.

**CHAIR**—Would you put Canada in with those two countries?

**Mrs Cameron**—I would not because we have just had a major teleconference with the National Library of Canada and they are not addressing the preservation issues at the moment. Their approach is quite different. There is quite a lot of online Canadian government publishing. They are working quite closely with Canadian government departments and they are mainly capturing government material only, whereas we at the moment are capturing very little because most of it still exists in print. What they do is only capture what we call flat HTML or text files. Because of the sorts of publications we are capturing, we have lots of what we call bells and whistles in our publications-lots of software plug-ins, audio and video. For example, we have a lovely Vegemite site where little characters jump up and down. It is not so much an issue of actually capturing that software, it is how we will migrate it in the future; how we will move it to a new platform.

**Ms Fullerton**—And maintain the look and the feel and the content.

**Mrs Cameron**—Yes, so that we keep that audio and video authentically, if you like. That is a major issue for us.

**CHAIR**—If any members ask to, let them come in because it is an informal briefing. I do not wish to interrupt your presentation. If you want to defer a question, just say so.

**Mr HARDGRAVE**—Going back to that cultural sensitivity matter in relation to certain Aboriginal groups, there have been reports in recent days that an Austrian tourist confessed that she was in the midst of North Queensland and picked up a bit of spare work painting traditional Aboriginal art for sale to the tourists. Have you satisfied yourself that perhaps part of their concern may have been lack of copyright-that, by our measure, copyright is probably something that protects them? If their images were on the Internet, the Austrian tourists would not even have to come to Australia to start doing the copy work-they could do it offline, couldn't they? What work do you do to try and satisfy groups that you may get images from that their rights are protected?

**Mrs Cameron**—All we can do is make sure that on the entry page-the page you have to go through before you actually get to the publication-we have a standard copyright warning that the National Library has written. We also have a link to the copyright warning that is usually contained within the publication. But we cannot absolutely guarantee that a user coming into the site will observe the copyright provisions. There is nothing we can do.

**Ms Fullerton**—They would be no more under threat on our site than they would be on the original site.

**CHAIR**—I was wondering whether you would publish material that was not covered by a copyright, if you see what I mean. Would you satisfy yourself as to who owned the rights to the material?

**Ms Fullerton**—Indeed. Once we have something in our collection, we observe the copyright laws.

**Mr HARDGRAVE**—Yes, of course you would.

**Mrs Cameron**—That is caught up with the fact that we would never capture a site unless we had permission and we had discussed the Copyright Act.

**CHAIR**—Colin, I think you had a formal presentation, too. I have to keep in mind our target of 8.20 a.m. Are there any more questions on that issue?

**Senator LUDWIG**—As I understand it, you capture Internet sites with copyright. What do you do about the changing nature of the Internet? What happens if the site that you have captured in a line changes? Also, how do you verify that the person who has put the material on in the first place is the copyright owner, for argument's sake, or is an authentic person—that is, from an Aboriginal cultural group?

**Mrs Cameron**—At the moment we usually have email or phone contact. We accept what we are told by the creator of that site. Essentially, they have to give us the Internet address and that is what we capture. It is probably a little difficult for somebody to forge something like that, if you know what I mean. It is not a particularly complex process. I doubt if somebody could really claim to be the owner and creator and use the address.

**Senator LUDWIG**—I am asking: what happens if someone sets up an Internet site which has cultural significance for a particular race and that person has nothing to do with the race—they are simply promoting it?

**Mrs Cameron**—No, we do not. We do not do so in relation to print publications or other information either. If somebody publishes a book about something or other and says it is whatever it is, we get it on legal deposit; we do not try to authenticate it.

**Ms Fullerton**—We are completely dispassionate. We collect all points of view, all things.

**Mr HARDGRAVE**—But the great fakes and forgeries are often accidentally captured and shown as originals anyway. That happens on occasion.

**Mrs Cameron**—In relation to the time line, we have set up capture schedules where for most publications we need to go back on a quarterly, monthly, weekly, six-monthly or whatever basis. That does add huge complexity to the operation.

**CHAIR**—I am going to flick over to Colin. Jasmine, I am sorry to cut you off, but I have a plan in mind to cover the things we all want to ask you, even though this meeting will have to end shortly. We are going to write to you.

**Mr Webb**—It is generally recognised that there are two major preservation hurdles that we have to overcome. One is the instability of the media, the things that the data is carried on. Things like floppy disks and CD-ROMs have a fairly short life span compared to things like paper. The more pressing hurdle is the obsolescence of the technology. As the technology changes, we can no longer read the data, even if it is on a permanent carrier of some kind.

This means that for preservation purposes we are moving away from thinking about preserving the object and looking for archival processes. So we are not looking for archival products any more. We are no longer looking for something that will last for 1,000 years; we are

---

looking for processes that will maintain accessibility. In looking for those solutions, generally speaking there are three approaches that have been adopted. One is to say, 'Let's keep the technology as long as we can. Let's keep the computer. Let's keep the software, so that when the—

**Mr SIDEBOTTOM**—You are a library and a museum.

**Mr Webb**—That is right. People generally accept that that has not got a future. It may be a bridge for a little while until we solve some of the problems, but long term it is not going to work.

Of the other two approaches, one has been mentioned here, which is migration. That is translating what you have into another platform that works in the future. There is a history of doing that for what we call flat formats—things that are fairly straightforward. A lot of the material is going to be very difficult to migrate, and we do not know how to do that properly yet, but we still think that is probably going to be the way of the future. The other approach people talk about is called emulation, where you develop software which will make your publication or your information work the way it used to work. You are not migrating the information forward; you are basically migrating the tools forward.

**CHAIR**—Sort of like a time machine.

**Mr Webb**—That is right.

**Dr Cathro**—A modern computer emulating an older computer.

**CHAIR**—I can just see a TV series developing out of this.

**Mr Webb**—We do not know how to do those things yet. We know how to do parts of them. We think the solutions to those problems have to be derived collaboratively. We are working with people all over the world, trying to identify people who are working or have an interest in that area, just sharing information and ideas. There is a large research effort going on. As Jasmine said, we are probably in the forefront of it, or in the vanguard anyway with some other people.

**CHAIR**—Colin, thank you. Because of the time constraint, senators and members will have to leave and we have to finish our formal talks. There are still a lot of questions that we would like to ask. Your presentation has given us a lot of insight and left a lot of unanswered questions. Because of the time constraint, if you and the members are happy to do so, I suggest that Lexia write to you with a number of questions that members would like to have asked—we will have another discussion about that—and you give us an informal written response. We do not want a great textbook of a response because this is preliminary; and/or we will add on another discussion at a convenient time next year. We are very concerned that parliament at least be running alongside—helping, supporting and being a partner—so that we are not caught short and do not lose valuable public records and the like. I think we have got a partnership here of great mutuality.

**Ms Fullerton**—I think we do. We definitely need a national approach to deal with this, and at the moment the National Library is struggling. It is a very labour-intensive and costly business. At this stage, there has been no allocation of funds anywhere in all of Australia specifically to do with this.

**CHAIR**—That is a very significant statement. I think members will appreciate the role this committee might have. Thank you very much for your assistance and cooperation, and may I also wish you a happy Christmas. Thank you for the good work you are doing.

**Ms Fullerton**—I would like to issue an invitation to you to come, singly or in groups, and have a look at some of these things that we are doing, because the impact from actually seeing them is much greater than that from just talking about it.

**CHAIR**—Yes, we will take you up on that, Jan. Thank you.

**Committee adjourned at 8.24 a.m.**