

Preserving the web evolution: AUSTRALIA'S PANDORA



All over the world, major Libraries and other national organizations are responding to this challenge - the potential loss of information - by collecting, preserving and giving permanent access to key websites for future generations.

PANDORA, Australia's Web Archive, is a project by National Library of Australia in conjunction with nine other participant organizations from all over Australia's States, and it is actually a selective archive of significant Australian online publications and web sites considered to be of long-term research value. It is based at the National Library of Australia in Canberra.

The National Library aims to archive titles of national significance, while the State Libraries aim to archive those of State and regional significance. Then the National Film and Sound Archive takes responsibility for sites relating to music and film; the Australian War Memorial archives sites relating to Australian military history; and the Australian Institute of Aboriginal and Torres Strait Islander Studies archives publications and web sites of Indigenous peoples.



**NATIONAL
LIBRARY
OF AUSTRALIA**

Work began on defining selection guidelines in January 1996, and the first titles were archived in October of that year. Nowadays, the archive is about 6 TB big, and it is constantly growing: but the significant element is that the Archive contains only a small proportion (estimated at less than one per cent) of the Australian web domain!

Actually, the possibilities for Web archiving are two:

the selective approach, like PANDORA's, basing on a selection of potentially interesting websites to be harvested
the whole domain snapshot approach, which means that the system automatically captures the entire web publications of a certain domain

There are advantages and disadvantages to both approaches, but the main difference seems to be in the actual functionality of whole domain snapshot: for example, there is no quality assurance applied to any of the resources (it would also be impossible to check the entire snapshot because of its sizes), and the result is that documents can be incomplete or lack functionality. It is also possible that the harvesting robot misses some resources, and - as the harvest is done without copyright permission from publishers - that some restricted websites may not be gathered.

On the other hand, from a Library's point of view, it is nevertheless a pity that the entire domain would not be collected, therefore the ideal situation would be a selective archive supplemented by periodic snapshots of the entire domain. The National Library undertook a harvest of the entire Australian web domain for six weeks during June and July 2005. In other countries, libraries are adopting hybrid approaches to develop a suitable solution for this challenging matter.



PANDORA contains a wide range of publications and Web sites, about Australia or regarding a subject of social, political, cultural,

religious, scientific or economic significance and relevance to Australia, and written by an Australian author. High priority is placed on collecting government publications and academic e-journals. In addition there are many categories to collect the other sites, as for example: Cultural activity, Community concerns, Scientific standards and research, Politics and government, Indigenous peoples, Sport, People (sites of well-known as well as 'ordinary' Australians).

The applied software, called PANDORA Digital Archiving System (shortly PANDAS) has been developed by the National Library of Australia to support its work flows, processes and metadata requirements, as well as those of other participants, in building the selective web archive. It is web-based software that enables collection managers in various geographic locations to contribute to a central archive.

Anyone in the world with an Internet connection can have access to the PANDORA Archive, and the system is very user-friendly and simple to surf.

To visit PANDORA please go to <http://pandora.nla.gov.au/>