The J S Battye Library of West Australian History

Guidelines for the Selection of

Electronic Publications

for the PANDORA Archive

October 2002

Contents

1 Introduction

1.1 The J S Battye Library of West Australian History recognises that there is a need for action to archive and preserve the significant portion of Western Australia's documentary heritage that is being produced online. The Battye Library further recognises that the State collection should form part of a nationally accessible collection.

1.2 The publicly available National Collection of Australian Online Publications currently consists of items archived in the PANDORA Archive by the National Library of Australia and its partners.

1.3 The Battye Library has responsibility for selecting, cataloguing and archiving items with a Western Australian coverage or focus as part of the PANDORA Archive.

1.4 The following guidelines outline the basic criteria for use by the Battye Library in selecting items for archiving as part of the National Collection of Australian Online Publications.

1.5 The guidelines cover publications located on the World Wide Web, at gopher and ftp sites or distributed via email. They do not cover databases,[1] information search services, discussion lists, bulletin boards and news groups.

2 Terminology

The following definitions apply to terms used in this document.

2.1 'Online' is used in preference to 'electronic', 'networked' or 'digital' when describing documents published on the Internet.

2.2 'Home page' is used in the limited sense of the first screen or entry point for a site. It is the default page that the reader is taken back to when s/he clicks on the home page icon within a site.

2.3 'Site' is used to mean a collection of linked documents, mostly with the same basic Internet address (internal links), although there are often links to documents on other sites (external links) as well. It is a general term, which includes serials, monographs and what are often referred to loosely as 'home pages'.

2.4 'Title' is used in the sense of an entity that is being considered, or has been selected, for collecting. It may be a single document, or a number of linked documents with the same basic Internet address. For discussion on defining title parameters, please see section 4.5.

2.5 In these guidelines, the term ' to archive' is used to mean the act of downloading from the Internet and storing on the National Library's PANDORA server. The term 'to preserve' is used to mean taking certain steps, including archiving, organising, describing, refreshing and migrating of titles, to ensure that the Battye Library can continue to provide long-term access.

2.6 For discussion of 'publication', please see paragraph 3.1.

3 Selection Intentions

3.1 The agencies contributing to the National Collection of Australian Online Publications operate on the basis that anything that is publicly available on the Internet is published. However, distinctions between traditional categories of documents such as books, serials, manuscripts, working drafts and organisational records are blurred in the electronic environment. It is not the intention to preserve organisational records and similar materials, which are the domain of archives and record management.

3.2   The Battye Library has separate guidelines for the collection of resources in print format. Online publishing is different from print publishing and will be treated differently. Publications are often mounted on the Internet without the quality filtering mechanism provided by editors and publishers. Consequently, standards are much more variable than they are in print.

3.3   The potential volume of online published material to be dealt with is overwhelming in relation to the resources available and proportionally more effort is required to access and collect them. A higher degree of selectivity will therefore be applied for online publications than is the case with print

3.4   The Battye Library will not attempt to preserve all versions/editions of a selected online title, just as we do not attempt to preserve all stages of a print loose-leaf item. In the online environment, publications can and often do change frequently and it is not feasible to capture all instances of change. A decision will be made on the frequency of capture desirable for each title. The publication pattern, the importance of the information, and the stability of the site will be factors that affect this decision. Some titles will be captured as comprehensively as possible, while others may have just a one off 'snapshot' taken of them.

4   General Selection Guidelines

The following selection guidelines are to be interpreted within the proviso that a higher degree of selectivity will be applied for online publications than is the case for print.

4.1   *Western Australian content*

4.1.1   To be selected for preservation, a significant proportion of an online work should

- be about Western Australia or be written by a Western Australian author;

- be on a subject of social, political, cultural, religious, scientific or economic significance and relevance to Western Australia. Some subject areas may be identified for more intensive collecting.

4.1.2   It may be located on either an Australian or an overseas server. Western Australian authorship or editorship alone is insufficient grounds for selection, except in the case of literary works (see paragraph 5.8). Content is the pre-eminent factor.

4.2   *Multiple versions*

4.2.1   In general, where there are both online and print or microform versions of a publication available, the print or microform will be acquired in preference to the online version. The online version of a print publication will usually only be selected if it has significant additional information or value.

4.2.2   Where there are both online and CD-ROM or floppy disk versions available, the online version is normally preferred. However, when there are effective differences in content and/or presentation both versions may be sought. In cases where, for technical reasons, essential content cannot be downloaded and/or useful functionality retained, the physical format version may be sought either instead of or in addition to the online version.

4.3   *Authority and research value*

4.3.1   High priority is given to authoritative[2] online publications with a high likelihood of future use or long-term research value.[3] Where an online publication provides a superficial view of information that is readily available elsewhere, either electronically or in print, it will generally not be selected for preservation.

4.4 Social and topical issues

4.4.1   In addition to authoritative publications, examples of other online publications which provide a general insight into Western Australian society and culture, and the way the Internet is being used by Western Australians, may be selected on a limited basis. More inclusive selection guidelines will be applied to online publications on social and topical issues of specific Western Australian interest (e.g. online ephemera such as festivals and other events), or to support particular collecting strengths of the Battye Library in other formats. The intention is not to duplicate the print collections, but to complement them by providing the broader context.

4.4.2   Sites for selected events or on particular subjects may be sampled during a limited time period and gathered together in a collective entry on the PANDORA Archive (e.g. '2001 Western Australian state election campaign').

4.5 Defining title parameters

4.5.1   Both higher and lower links on the site are explored to establish which components form a title that stands on its own for the purposes of preservation and cataloguing. Internal links only are archived.

4.5.2   Preference is given to breaking down large sites into component titles and selecting those which meet the guidelines. However, sometimes the components of larger publications or sites do not stand well on their own but together do form a valuable source of information. In this case, if it fits the guidelines, the site will be selected for archiving as an entity.

5   Selection Guidelines for Specific Categories of Material

5.1 Annual Reports

5.1.1   Annual reports from significant Western Australian-based companies or organisations may be selected if they are not available in print.

5.2 Government Publications

5.2.1   The Battye Library accepts primary responsibility for preservation of Western Australian Government publications only available online. State Government publications will be collected comprehensively. The distinction between publications, organisational records and working drafts is sometimes difficult to determine in the online format. It is not the intention of the Battye Library to archive documents in the last two categories.

5.3 Digitised Materials and Online Exhibitions

5.3.1   To be considered, the site must contain more than digitised copies of items, and include, for example, historical or interpretative information or other intellectual input.

5.3.2   In general, sites containing materials such as manuscripts, archival records and pictures that have been digitised by government-funded collecting agencies, will not be selected. Sites made available by other organisations will be considered on a case-by-case basis.

5.3.3   Online exhibitions mounted by Western Australian organisations and on topics of importance to the State's history or culture may be selected if there is no printed catalogue available, or if the online exhibition demonstrates significant additional features. To be considered, the site must contain more than digitised copies of items, and include, for example, historical or interpretative information or other intellectual input.

5.4 Educational Material

5.4.1 Sites created for educational purposes will not be routinely selected. However, some examples of Western Australian-based educational material will be selected if they are of particular quality in terms of presentation or content.

5.4.2 Award winning Western Australian school sites may be archived from time to time as representative examples in this field.

5.5 Ethnic Community Sites

5.6.1 Ethnic community organisations' sites will be selected if the focus is information about the experiences, activities and concerns of the community in Western Australia.

5.6 Juvenile Publications

5.7.1 Sites created for or by juveniles will be selected on a limited basis. The intention is to sample sites which demonstrate the use of the Internet by young people and which represent their concerns, attitudes, lifestyle and culture.

5.7 Literary Works

5.8.1 Western Australian literary works (e.g. novels, poetry or short stories) will be considered for archiving, whether or not the author is previously established in print.

5.8 Newspapers

5.9.1 Sites for newspapers that mainly duplicate the information provided in print will not be selected for preservation.

5.9.2 Western Australian newspapers available online only will be assessed against the guidelines and preserved if they meet standards described in Section 4, including authority, quality and originality of content.

5.9 Organisational Sites

5.9.1 Western Australia organisational sites will be selected if they contain substantial information not available elsewhere (e.g. in annual reports).

5.10 Personal Sites

5.10.1 Personal sites will be selected on a very limited basis if they provide information of research value unavailable elsewhere.

5.11 Sensitive Materials

5.11.1 Sensitive or legally questionable online material deemed to have political or cultural significance for Western Australia may be selected, but access restricted to designated researchers.

5.12 Subject Specialities

5.12.1 The Battye Library, from time to time, may determine specific subject areas to focus on. These target areas will be publicised in annual plans.

5.13 Exclusions

5.13.1 Drafts and works in progress will not be selected, even if they otherwise meet the selection guidelines. Wherever possible, completed documents only will be preserved. (This point does not refer to unpublished literary manuscripts. The Battye Library may well wish to negotiate with certain authors regarding the preservation of drafts in electronic format of literary works in progress.)

5.13.2 Items deemed not to be publicly available, e.g. material on Intranets, commercial-in-confidence material.

5.13.3 Promotional sites and advertising (occasional samples relevant to Western Australia may be taken and gathered together in a collective entry).

5.13.4 Directory/portal sites that only serve the purpose of organising or linking to Internet information.

5.13.5 Organisational records.

5.13.6 Theses made available on the Internet. These are the responsibility of the universities and the Australian Digital Theses Project led by the University of New South Wales.

6   Review

These guidelines are reviewed on an ongoing basis.

For further information please contact:

Brian Stewart
Email: bstewart@liswa.wa.gov.au
Telephone: (08) 9427 3300

7   Footnotes

---

(1) This category includes commercial dial-up databases, library catalogues, statistical databases, geographic information systems and other large databases requiring search engines. It does not refer to publications which happen to be published in a database format, even though a search engine may be employed.

(2) In order to determine the authority of a title, it should be clear which organisation or individual is responsible for producing it and what the qualifications or expertise of the creator/s are. Ideally, background information about the organisation or individual should be available. It may be necessary to research the legitimacy of the organisation, the individual or the information. There are a number of sources on the World Wide Web regarding the evaluation of online publications. One is the Widener University/Wolfgram Memorial Library *Teaching Critical Evaluation Skills for World Wide Web Resources*, http://www2.widener.edu/Wolfgram-Memorial-Library/webeval.htm. This site contains a bibliography of Web evaluation techniques.

(3) In relation to print materials in libraries, 'research value' has been defined very broadly. For example, even a single sheet of junk mail in the ephemera collection can be considered to have research value, in the sense of revealing information about an aspect of society at the time of its publication. In the context of online publications, 'research value' is much more narrowly defined and means that the publication should be a substantial compilation of information.