Virtual Preservation: How Has Digital Culture Influenced Our Ideas about Permanence? Changing Practice in a National Legal Deposit Library

Ingrid Mason

Abstract

This two-part article considers how digital culture has influenced ideas about permanence. It examines the change in collecting practices in one legal deposit library. The author considers how the idea of permanence, understood in cultural heritage terms, influences digital culture, and, thus, digital technology. The first part of the article addresses the concepts associated with permanence, digital culture, digital technology, social change, and cultural institutions, in relation to collecting digital cultural material. The second part focuses on changing collecting practices of the Alexander Turnbull Library at the National Library of New Zealand for electronically published material with the benefit of legal deposit.

Outline

The first part of this article considers the concepts associated with permanence, digital culture, digital technology, social change, and cultural institutions, in relation to collecting digital cultural material. This is intended to place the change in collecting practices, outlined in the second part of the article, in the context of an evolving understanding of how these concepts might be interpreted and are being applied. The second part focuses on the change in collecting practices of the Alexander Turnbull Library (Turnbull Library) as it develops its heritage collection of electronically published material with the benefit of legal deposit,[1] with particular attention to the change in practice to include the collection of online publications.

CONCEPTS

This *Library Trends* issue presents preservation in cultural heritage as its broad theme, and this section questions specifically the influence of digital culture on ideas of permanence. Implicit in the question "How has digital culture influenced our ideas about permanence?" is the assumption that digital culture has had, or is having, an influence upon ideas of permanence. But, is that true? Answering this would require exploration in greater depth than is possible in this article, but it is possible to offer up institutional practice as a means of responding to the question. Another question needs to be asked: how is the idea of permanence, understood in cultural heritage terms, influencing digital culture and, thus, digital technology?[2]

Digital culture is expressed through social, cultural, political, and economic activities that are undertaken using digital technologies. The presence of digital technology and the centrality of its use distinguish these practices and activities from practices and activities that are undertaken using analog technologies or no technologies at all. Ideas of retaining and restoring culture, authenticity, and the regular reexamination and reinterpretation of culture are heavily threaded through cultural heritage discourse, heritage legislation, and institutional policy. People continue to want cultural material collected, looked after and made accessible, whether it is analog or digital. Research interest in digitized heritage material and increased institutional commitment to digitize analog material reflects a link between the demands of digital culture for online access to digital heritage material, and the force of continuing interest in the past—clearly seen in the rise in online (and remote) genealogical research at most cultural institutions. But the nature of digital culture, the material difference of digital cultural heritage, the increasing volumes of digital material produced, and expectations of access and online availability have an impact on notions of collecting: notions such as collecting everything; keeping everything in the manner in which material has been kept before; digital material as original, untransformed and complete; methods and technologies used for acquiring and preserving digital material; and modes and technologies used to access digital material.

Digital technologies have an attendant hype of panaceas or apocalypses. They offer faster computing power, faster rates of update or change, different types of interactive and immersive experiences to that of analog technologies, and they stimulate an interest in what is new, or what is possible, rather than what was. They engender a pressure to respond to intensified rates of change and higher levels of attrition or loss of digital material, and a need to ascertain where or how human oversight and intervention is most feasibly applied to capture what was, and to prepare for what is new and what is possible when collecting digital material.

Digital technologies enable continual change and improvement to

processes and outputs, through the deployment of novelty.[3] The impelling nature of technological innovation creates two significant complexities from the perspectives of digital acquisition and preservation. The first is the organizational resources and processes that are required to anticipate and respond to the rate of change. Technological innovation per se is unpredictable and volatile, and, in itself, poses feasibility issues for collecting organizations and their fitness to respond proactively to develop the means to acquire and preserve digital material. The second is the opportunity and the right to grapple with the technical implications of change. Proprietary technological innovation tends to develop proprietary formats and applications, posing legal and efficiency issues for collecting organizations, and reducing their ability to openly examine file formats and applications and, thereby, to develop stable collection and preservation strategies. The debate over opening up the documentation of RAW digital image format and the challenge to digital camera producers informs this issue.[4]

The digital technology development industry provides the means to go forward, as the dynamic of digital technology and digital culture demands. The cultural heritage sector issues an equally forceful challenge, driven by continued public interest in cultural material, for technology to be developed that enables people to go forward and backward easily, and to retain the same access to digital content and the experience of accessing it "as it was." Flexibility that enables digital collecting and preservation to progress in such a volatile environment needs to be built into digital technology. The spiral development referred to by Mackenzie Smith (2005) for the digital archive at MIT emphasizes this point; stability, too, needs to be built into digital technology to permit long-term collection, preservation and access and, thereby, to enable long-term research using digital cultural heritage material.[5] The development industry has yet to take up the challenge of providing the means and flexibility to go backward—a requirement common to collecting institutions and to consumers.

The idea of permanence, as it is understood in the cultural heritage field, is asserting itself upon digital culture and technological development, just as much as digital culture and technology is asserting its requirement for greater flexibility in cultural heritage practice. People have a continuing need to go backwards with ease and "mark the spot," or experience accessing material in its "time" digitally (e.g., to cite a journal article in an academic paper by linking through a permanent identifier to an online journal, or to play a computer game developed to run on Windows 3.1 in that operating environment, or in one that emulates it). Research and cultural interest in historical cultural content (digital and analog) has not waned; it is evident in the development of permanent identifiers and of emulation technologies. Recent research at the National Library of Australia indicates that the Web sites mostly frequently

used in its PANDORA archive are usually those that are no longer available on the Internet (Crook, 2006). Metadata standards, such as those for recordkeeping metadata for Australian government archives (National Archives of Australia, 1999) and preservation metadata outlined by the PREMIS Working Group (2005), and preservation strategies such as file format migration and emulation, are a response to a cultural demand for permanence in digital terms.

The need to fix things in time and retain artifactual and documentary material from the past is to a degree forensic in nature; authenticity is crucial to society's understanding of historicity, whether measured in terms of centuries or seconds. Pressure is being asserted on digital technology to meet these interests and needs, so that questions, such as "what happened?" or "what was?," can be answered with a degree of confidence—confidence that the evidence being examined, or material utilized, is as consistent as possible with what was examinable when it was created and used, and not been altered to skew its content or context and, thus, its potential meaning. Digital culture wants to continually revise its past as much as project into its future; digital technology will need to evolve to meet and satisfy that desire.

At what level can the word *permanence* be applied to cultural heritage practices? Permanence is a vital principle of cultural heritage: the raison d'être of collecting is to retain a cultural identity and to build up the resources—the cultural and research collections—that permit cultural enrichment, facilitate research, and bring wider social and economic benefits to the society that supports and finances that collecting activity. In principle, permanence is key, and, to a great extent, permanence is key in practice too, in that the business—the operations, sourcing, selecting, acquiring, preserving, and making available material—remains constant. In cultural collecting, permanence applies to *why* cultural material is collected. It is, however, in determining *what* that cultural material is and *how* that business is undertaken that changes in practice are taking place. Anticipating and meeting the needs of researchers, developing digital collections and addressing issues of digital preservation remain a considerable challenge; there are many unknowns in establishing new practices for collecting electronic publications.

Social change resulting from the emergence of digital culture is affecting the operational practices and procedures associated with collecting and preserving cultural heritage at the Turnbull Library. Cultural institutions, such as the Turnbull Library, are also social institutions, and the tensions associated with steering a steady and relevant course in times of rapid social change are not new. Cultural information and knowledge is accrued by cultural institutions and professionals all the time and over time; their understanding and practices are used to develop, maintain, and provide access to heritage collections. Cultural practices are embed-

ded in the development of a cultural institution's collections, organizational processes, systems, culture, and people, and in the relationship they have with the community. Cultural information and knowledge is fed out into the community and back to the institution. Metaphorically speaking, cultural institutions are in the business of slowly crafting and shaping social and cultural fabric.

Cultural institutions need to be robust enough to absorb the uncertain and complex aspects of social and cultural change, and yet fluid enough to evolve correspondingly to support and present this change. But there is a tension between such fluidity and fixity, which, as Brown and Duguid (2000) note, serves an equally important purpose. Fixity gives a sense of direction:

> There are good cultural reasons to worry about the emphasis on fluidity at the price of fixity. But fixity serves other purposes. As we have tried to indicate, it frames information. The way a writer and publisher physically present information, relying on resources outside the information itself, conveys to the reader much more than information alone. Context not only gives people what to read, it tells them how to read, where to read, what it means, what it's worth, and why it matters. (Brown & Duguid, 2000, p. 201)

This links directly to the role of cultural institutions, which provide a sense of the past, present, and future—cultural and social, fixity and fluidity—on a continuum, irrespective of technology.

It is important to acknowledge these inherent tensions in any response to digital culture. The rate of change, the volume of digital material being published, and the diversity of digital technology and digital culture overwhelm the possibility of applying the same level of human intervention as with analog practice. It is no longer possible to maintain the level of manual processing and to achieve the same levels of comprehensiveness in collecting, and digital preservation methods are nascent. New methods and approaches to managing increasing publication production levels and technological innovation, and a redefinition of acceptable levels of collecting to retain the corpus of electronic publications of a nation, are being developed and implemented.

Changes in Practice

The Turnbull Library develops its heritage collection of published material under legal deposit provisions of the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003. This recently amended legislation has extended the Turnbull Library's collecting reach to electronic publications. The legislation defines public documents as those "printed or produced by any other means in New Zealand, or is commissioned to be printed or otherwise produced outside New Zealand by a person who is resident in New Zealand or whose principal place of business is

in New Zealand" (National Library of New Zealand [Te Puna Mātauranga o Aotearoa) Act 2003, s29[1][b]). Thus, electronic publications distributed either as offline publications (made available in portable format) or online publications (made available via the Internet) come within the ambit of the legal deposit.

The Turnbull Library's collections of published materials include monographs, serials, cartographic and audiovisual materials, special print and rare books, and ephemera.[6] Before the change in legislation, offline and online New Zealand electronic publications were acquired through purchase and by permission. Acquisition of offline and online New Zealand publications is now covered by legal deposit. The intent to collect New Zealand publications comprehensively is consistent with the library's legal mandate, which remains unchanged.[7] New, though, are the types of publications being collected by the Turnbull Library and how the intent to collect comprehensively is being realized. Collecting and keeping electronic publications has meant revisiting the principles that guide the practice of collecting publications at the Turnbull Library, and applying these principles to the collection of electronic publications.

To collect electronic publications as they are now produced, national and international publishing trends must be understood and monitored continuously to enable planning.[8] The Turnbull Library collects New Zealand publications, works published overseas by and about New Zealand and New Zealanders, and publications that relate to the Pacific and Antarctica. When harvesting material that is not published or located within New Zealand, permission is sought to collect, preserve, and make this material accessible (with the attendant rights observed). The inability to collect published material exhaustively and the need to consider the selection criteria for such material, which standards to follow or set, and which tools and processes can be developed to enable collection, are the key issues facing institutions like the Turnbull Library, whose mission is to collect the national corpus of publications irrespective of format (Illien, 2006; Masanès, 2005). The curatorial intent at the Turnbull Library is to forge a collecting approach to electronic publications that has links to its approach to acquiring print publications, particularly those analogous to traditional print forms, that recognizes readily new publication types as they emerge, and is willing to determine research value and consider what it may take to acquire these new publication types. Diverse curatorial approaches are being undertaken in other national libraries around the world, including voluntary deposit, collaborative selective harvesting, subdomain and whole-domain harvesting, and bulk transfer of digital material (National Library of Australia, 2004). All of these inform the Turnbull Library's curatorial decision making.

The National Library of New Zealand strategy that enables the Turnbull Library to collect electronic publications comprehensively is to em-

ploy diverse means to build a collection of electronic publications. Material is acquired for the collections through "push" and "pull" business processes. Publishers can "push" offline publications (on portable formats such as floppy disc, minidisk, CDROM, DVD, or hard drive) mostly through the post, and online publications through email or an electronic drop box, to the library, although the deposit of online publications is not required by legislation. The "pull" method currently involves Turnbull Library selection staff running Web crawling software to harvest Web material selectively. The Turnbull Library will soon also undertake domain harvesting.

With regard to the "push" business processes, the legal deposit provisions of the National Library Act require publishers to submit two copies of offline publications to the National Library of New Zealand. One of these comes to the Turnbull Library to keep in perpetuity in its heritage collection. As in the past with printed materials, publishers will be obliged to submit these publications in the portable formats they are published on. Legal deposit staff have been consulting with publishers of print and electronic publications before, during, and since the change in legislation. In August 2006, the gazetted requirements were enacted (National Library of New Zealand, 2006) and the legal deposit staff are now building working relationships with publishers newly covered by the legislation, and establishing deposit arrangements for both print and electronic publications. As publishers are not legally required to deposit online publications, legal deposit staff will seek publishers' assistance in depositing them electronically. Government publishers (central and local) and tertiary education publishers are being approached first. This approach mirrors the workflow for print monographs and serials, for example those produced as PDF, Word or RTF documents, and the electronic output of these publishers often have high research value and may well be missed in the periodic harvesting that will be undertaken.

Selective and domain harvesting is being undertaken because of the rich research value found in material that has not been published traditionally but is now available on the Web. Small-sized selective harvests based on subjects, themes, and events are being undertaken. The Turnbull Library's selective harvesting draws upon the curatorial approach to selective harvesting undertaken by the PANDORA (http://pandora.nla .gov.au/index.html) and UKWAC Web archiving consortia (http://www .webarchive.org.uk). Larger-sized harvests based on sub-domain (defined as .govt, or .org or .co/.com, etc, within the larger whole top-level domain) or whole-domain (that is, all Web sites registered in New Zealand and New Zealand Web sites registered outside of New Zealand, including .nz as the country code and .com or .org as the top level domain) are yet to be undertaken.

New tools and technologies are being employed to enable the collec-

tion of electronic publications. An online submission tool has been developed so that publishers can upload published material for selection by legal deposit staff, using selection guidelines developed by the Turnbull selection staff. Domain harvesting is on the agenda for 2006/2007. The means to do this have yet to be identified and applied. The National Library of New Zealand will also look at bulk upload or transfer of digital material and the deposit of databases and data sets. The databases in the deep Web have been identified as rich deposits of published information (Bergman, 2001) and the Turnbull Library is interested in acquiring this type of electronic publication. It is important to maintain awareness of the endeavors of other cultural institutions, publisher interests, any changes in publishing technologies and production patterns and compliance regimes for publishers. While it is desirable to extend the methods of acquiring digital material, resourcing requirements and capacity will continue to be constantly monitored and evaluated to ensure that their efficiency and effectiveness. However the material is acquired, once it is acquired, it is then destined for the National Library's National Digital Heritage Archive, when it is implemented. Material that is currently stored in the Object Management System, the National Library's interim digital repository, will be transferred to the National Digital Heritage Archive.

Turnbull Library's harvesting tool changed in 2006. Thanks to considerable consultation and support from members of the International Internet Preservation Consortium (IIPC), the National Library of New Zealand and the British Library have together developed the open source Web Curator Tool (https://sourceforge.net/projects/webcurator), to manage the selection, acquisition, and appraisal workflow of selective harvesting. The WCT was implemented in October 2006 at the National Library of New Zealand. The cultural heritage field is renowned for its collaborative work and interest in efficiency, and this is a good example of the pooling of skills, resources, and expertise that enables the realization of new initiatives with shared interests and benefits. Other collecting organizations can use the WCT to harvest Web material, contribute to its enhancements, and provide insights to other curatorial and technical practices in building professional knowledge in this arena.[9] Networking with peers has been vital to validate or contradict experience, to debate and challenge traditions and perceptions, and to lead the change behind the scenes well before it is reported in cultural heritage discourse. The work at the Turnbull Library has benefited from the insights of fellow practitioners dealing with electronic publications, in particular online publications, at the National Library of Australia, the British Library, the Wellcome Trust, the Library of Congress, Library and Archives Canada, and the State Libraries of Victoria and New South Wales in Australia.

Decisions made in selection, acquisition, appraisal, and preservation determine the presence and longevity of cultural heritage material.

Some electronic publications will inevitably not make it into the Turn-bull Library's electronic publications collection. Electronic publications may have already vanished or will vanish in the intervals between domain crawls, may not be selectively crawled, or may be rejected during appraisal of harvested material because of damage or loss during technical transfer. It is not feasible to retrieve, let alone acquire, some electronic publications (such as content lost or deleted in dynamic databases or residing on decaying portable format); nor is it possible presently to preserve some electronic publications, because of their unknown or unstable file formats. Concentrating on material of high research value that can be captured now, rather than being obsessive about what is missed, is our current strategy. As with any enterprise associated with value and risk assessment, it is important to be clear about the principles, processes, and priorities driving the activity, and to keep the variables in perspective; not everything can be done at once, and not everything will be perfect. An example of pragmatism driving change in business process is the recognition that publishers and researchers benefit from having the publishing community deposit "traditional" types of electronic publications with rich research value, that is, those produced in simple formats such as PDF and Word, as they publish them. This replicates the process undertaken in print, and connects readily with processes already in place for acquisition and cataloging. The benefits to publishers of getting their publications cataloged and listed in the national bibliographies are well established as a means of driving sales.

The diverse information architectures, technologies employed, and content embedded in Web sites, pose challenges for harvesting. This is the case for selective harvesting in particular, which is driven by an intent to capture material of high research value and therefore focuses more intently on a deeper harvest of a Web site than domain harvesting offers. There are common practices in all of these areas of Web design and production, but there are no enforced standards to aid with analyzing Web site content and structure and configuring harvesting settings accordingly. Selectors and Web archivists need knowledge of Web design and construction, rates of content change, and production trends to assist them in their decision making for selecting material and scoping harvests to capture Web materials. For example, Web material can be closely examined and scoped for selective harvest. The harvester settings and schedules are applied to capture Web material in a manner befitting the research value of its content and its technological dynamism. The Web sites of registered political parties exemplify this. These are mostly unique, in that there is no print equivalent for most of their content. However, at different times and for different reasons, their collective significance may change. Year in and year out all of them have equal significance, but in an election year those of the leading parties may offer more significant content, or those

that engage on an issue of high political interest, such as welfare payments or environmental regulation, may acquire greater social significance and, thus, greater research value and be harvested more frequently. Selectors or Web archivists grapple with these variables and the reasons for acquiring online publications as they evaluate and identify rich content, and set the timing of selective harvests, and appraise harvested Web materials.[10]

To demonstrate the idea of responsiveness, during the latest budget rounds of the New Zealand Parliament Turnbull Library selection staff selected government Web sites and blogs for harvest with a view to capturing news and debate relating to the budget. Ironically, the 2006 budget was not controversial and the Web content and commentary captured was correspondingly lacking in information and interest. By comparison, the rich commentary captured in blogs during the 2005 parliamentary election in New Zealand was impressive, and efforts to capture it were well rewarded; there is absolutely no equivalent of this Web content published in print.

Large content-rich or intensively dynamic commercial Web sites, however, are not suited to selective or domain harvesting; they offer further technical and curatorial challenges.[11] New Zealand examples are *Te Ara: The Encyclopedia of New Zealand* (http://www.teara.govt.nz) and *TradeMe: New Zealand Online Auctions and Classifieds* (http://www.trademe.co.nz), both of which have research value. National encyclopedias have long been significant cultural and research publications in print form, and continue in their electronic forms. *TradeMe* offers different research value in that it reflects a decisive social shift to trading online, and the movement of advertising of new and second-hand goods from print to mostly electronic media. Institutions collecting cultural heritage have always responded to changes in society, politics, and technology, so this is nothing new. Simply put, new means are being established at the Turnbull Library to continue to achieve the same end—the collection of published documentary history.

With Web archiving in particular practices are evolving. Judgments about what is harvested and archived are being made now, but in two years' time they may be made differently. At present the Turnbull Library selection staff is undertaking selective harvesting based on topics of interest and expertise: music, ethnic communities, sport, arts, and crafts. These topics are far too broad to evaluate effectively the Web sites within them, so specific foci are applied to permit selective harvesting: organizations and recording labels (music); organizations and support resources (ethnic communities); rugby, netball, and golf (sport); and, crafts and craftspeople (art and crafts). Within these foci other guidelines for selective collecting apply: a comprehensive representation of national interests or activities, and a selective representation of interests and activities within the Wellington region, where the Turnbull Library is based, as a

priority. Selection staff have noticed how easy it is to select material when social structures, such as national or regional bodies, are well established, and where an activity has been continued over a long time. Selection in well-established and popular sports, such as rugby, netball, and golf reflect this. Where Web material is informal, less established, or created by individuals, selection is much harder, and subjective judgment is required to select in a representative manner. Selection in emerging, more fluid or specialized areas of society and social activity, such as recording labels, crafts and craftspeople, reflect this.

In the case of selective harvesting, supporting documentation helps set parameters to assist curatorial staff as they make decisions about the areas they are selecting in, and how they can approach their subject, topic, or event. These selection and appraisal decision-making templates have been designed to sit within a selection and appraisal decision-making framework; the templates and framework provide an intellectual structure for staff. The records of the decisions made by the curatorial staff provide an information base to refer back to in evaluating Web material for selection and for its retention, once harvested. These documents also form a foundation for curatorial understanding to guide the Turnbull Library's selective harvesting.

The selection and appraisal framework for selective harvesting at the Turnbull Library borrows heavily from archival theory. It places the decision making for selection and appraisal associated with selective harvesting in a collecting context and records the reasoning behind selectors' choices. Priorities for content areas can be driven by a selector's expert knowledge of the subject matter or by subject significance, or they may be related to other materials held in the collections. Entirely new forms of publication, new subject areas, or publishers will be added to the collection, expanding the range of the documentary forms or documenters already known and understood. In selective and domain harvesting, these methods of collecting are acknowledged to be representative. The curatorial approach to selective harvesting draws upon archival and museum curatorial practices, and an understanding of the need for representativeness in selective Web harvesting is building internationally. Research at the Bibliothèque nationale de France shows that selective harvesting permits a deeper crawl, whereas domain crawling permits a broader crawl (Masanès, 2005). For the Turnbull Library it makes sense to ensure that selective harvests are undertaken in a timely manner for material that has high research value, especially if those publications are more likely to disappear altogether.

The documentation of findings in appraisal work has been crucial in building up understanding. By recording and then synthesizing curatorial and technological observations, curatorial staff have collated evidence and developed the rationale that informs appraisal decisions. Two

new electronic publications' selectors at the Turnbull Library undertook this task after the selective event harvests of the 2005 New Zealand parliamentary election. It immersed the new practitioners in the harvested content, and allowed their instincts and questions to emerge in a relatively unconstrained way as they recorded their findings, which led to their appraisal recommendations. Not only did it become clear that material harvested from major political party Web sites and political blogs was extremely valuable, but it also became clear that content on smaller Web sites representing less popular political support or single issues did not change much over the period of harvest. Knowledge of the political and social issues, the controversies that arose, and the close election outcome all contributed to the assessment of the material. None of these discoveries seems particularly surprising—the proof, though, was most definitely in the pudding and it was an affirming exercise (Joe & Lala, 2006).

As noted, the Turnbull Library, with the help of other units of the National Library of New Zealand, especially the Innovation Centre and Bibliographic Services, has moved into the new business of acquiring electronic publications under legal deposit and is establishing feasible and acceptable practices. Different approaches to collecting electronic publications can be taken and they *all* have attendant benefits and risks. Prioritization for collecting and preservation can be undertaken in different ways with different rationales. For example, the earliest material may be prioritized for selection and preservation because it is less likely there will be documentation available for it or expertise to enable preservation to occur in the future. Then there is prioritization based on the uniqueness of material about which little is known and for which there is no equivalent or facsimile; or—the flipside—selection of material that is being produced now may be prioritized because it is easy to know and there is plenty of expertise around. Or, should one dive into the subjective area of collection assessment and put a research value on some material because it offers the most in terms of research return, determine what the "good material" is and select and try to preserve it first? (Cumings & Mason, 2004). Or, should one not make a subjective judgment and select what is technologically the most feasible to preserve at the outset, irrespective of what it is? Decision-making models are plentiful: Pareto analysis, cost-benefit analysis, decision-trees, etc. What has proved to be important is being able to fill these models with the information required to make good operational decisions, anticipating variables such as staff time, expertise and competencies, technology and project management costs, and social impact.

All of these decision-making scenarios offer reasonable outcomes, but they also present rather sticky ethical questions: how acceptable is the loss that occurs by omission, and which rationale has the most merit? A combination of these scenarios is one way of addressing the issue of

selection. Decision-making models and ideas are being developed to aid organizations that address these collection management issues (Woodyard-Robinson, 2006). Several event harvests have been conducted by the National Library of New Zealand: the 2002 America's Cup and the general elections in 1999, 2002, and 2005; major government agency Web sites have been harvested regularly since 2003. The rationale driving the Turnbull Library's selective harvesting has been based on staff time and competencies, technology availability, and research value. Similarly, the State Library of Victoria in Australia has established digital preservation procedures to guide decision making, and has designed digital preservation categories for items collected to prioritize digital preservation work. Simple questions are asked about the item's heritage significance, technological vulnerability, and scarcity (State Library of Victoria, 2005).

Libraries without a directive to maintain their research collections permanently are able to assess their collections and acquire, preserve, and dispose of research material in alignment with the needs of the funding body or community they serve. In contrast, the Turnbull Library maintains and makes available its collection material in *perpetuity*. In principle, all material acquired by the Turnbull Library benefits from that long-term investment. In practice, not only is it impossible to collect electronic publications exhaustively because of the sheer volume of material and the pace of technological change, but it is also impossible to acquire, preserve, and make electronic publications available perfectly. Nor should it be possible, as it has never been possible to achieve this with analog material, as attested by the challenges of preserving and providing access to fragile, degraded, volatile, large or non-standard format analog materials. It is unlikely that there will be sufficient resources to maintain digital material in its original form, unless its cultural and research value is equally high and there is a strong imperative to do so.

Diverse technologies and methods are employed, with others yet to be devised, to improve the Turnbull Library's ability to maintain and provide access to electronic publications. Efficiencies in manual handling and increased use of digital technology to do routine work are required if the Turnbull Library is to fulfill its mandate. In selective harvesting, several areas have already come under scrutiny for further workflow efficiencies where business process change and automation will assist: permissions (for example, the capacity to generate emails using data and templates in the harvesting tool to speed up workflow and enable responsiveness); quality review (for example, the capacity to tune the crawler to achieve more effective crawls resulting in less post-harvest fixing, and the capacity to visualize harvest results that would aid appraisal decision making); and description (for example, the capacity to automate cataloging, attribution of metadata, and/or full-text indexing to augment intellectual access).

The underlying premises of the operational changes are utilitarian or functional, but they are very clearly guided by curatorial principles and business efficiencies. It is important to value these items through acquisition, preservation, and description, but not to undermine the overriding principle—to retain cultural heritage—by attempting to do too much and failing to prioritize tasks and activities.

Digital culture has already exerted its influence on the practices of cultural institutions, such as legal deposit libraries. What, perhaps, are more interesting questions for cultural practitioners—aside from the current challenges of experimentation and implementation, the learning, the successes and failures in the response to the demands of digital culture—are: How are digital culture and digital technology going to respond to the demands of cultural institutions? What are digital users going to do when a cultural institution forces them to identify themselves online, as they would in a face-to-face situation, in an attempt to gain access to sensitive, privileged, or protected material? How will publishers respond to the interest in their material being selected under voluntary deposit or the legal requirement to comply?[12] Digital users are used to facing both open and gated material, and to accepting or subverting it as they see fit. Recent research shows that the generation immersed in the use of digital technology has very high expectations of getting access to vast amounts of digital material very quickly, if not freely, and of using it as they wish (Berkery, Noyes & Co., 2005). This provokes more questions: How will all the interests (of producers, collectors, and researchers) in digital material be balanced? How is digital material going to be collected? How will it be made available—freely or heavily constrained? Will all of those interests be satisfied equally?

Several forces are operating currently, including technocratic, individualistic, democratic, and commercial forces. It is the responsibility of cultural institutions to identify these forces, consider their institutional mandate, and respond, not necessarily with acquiescence, but with constructive, well-considered and planned action driven by their organizational intent and researchers' needs. In the case of the Turnbull Library, that intent is to collect comprehensively, while accepting that resources must be directed and used carefully, as it continues to collect, preserve, and make accessible its collections for the benefit of the community that it serves. The Turnbull Library must continue to gather and maintain the tangible and intangible value of published documentary history for New Zealanders in its collection of cultural heritage, analog and digital. Its practices are changing because practitioners are asking questions of themselves and of colleagues, experts, technology and digital communities, and are making informed choices. Some these choices are specific, for example, how to choose a Web site to harvest that has research value,

and how to go about harvesting it; other broader questions about preserving Web sites, whether they be simple and static or large, complex and dynamic, are yet to be answered.

So, how has digital culture influenced ideas about permanence at the Turnbull Library? It has certainly tested collecting principles and changed practice, has required a revision of collection policies, standards, procedures, and guidelines, and has stimulated change in business processes to enable the collection of electronic publications. It has provoked significant debate, and practitioners have had to reexamine what permanence means in operational terms when it comes to collecting, preserving, and making digital collections available. Certainly the modes and methods employed have some impact upon what is collected and retained, as do the resources available and the willingness to embrace change. Permanence is about being able to provide material in the collections and to support services that allow communities to trace ideas and events back in the past, draw them into the present, and project them into the future. There is a need to openly anticipate memory loss as much as memory retention, but what is not yet clear is what loss is acceptable and can be expected, and the impact that memory loss might or might not have (O'Hara et al., 2006). Whether that which is regularly used and enjoyed and of value to society now is prioritized for collecting and retention, in preference to that whose value is yet to be realized, or that which may have negligible value and, in fact, may never be retrieved, has yet to be resolved. These are contentious questions about the ethics of prioritizing preservation decisions, although this has long been the responsibility of curators and cultural institutions.

NOTES

1. Legal deposit in New Zealand supports the development of two collections: the Alexander Turnbull Library published collection, and the National Library of New Zealand general collection (see *National Library of New Zealand [Te Puna Mātauranga o Aotearoa] Act 2003*).
2. The works by Cameron & Kenderdine (2006), Phillips (2005), and Rabinovitz & Geil (2004) provide examples of practitioners reflecting upon the implications of their actions and decision making in their development and collection of digital cultural heritage.
3. An example of this is appcasting, a means of conveying software releases and updates through RSS (Really Simple Syndication) feeds (see http://connectedflow.com/appcast ing/).
4. More than two-thirds of 19,207 respondents to an international survey conducted by OpenRAW expressed concern about inability open or edit raw files created by older digital cameras. Ninety percent of respondents agreed:

    Once a digital image is written to a file by a camera, data in all parts of the image file should belong to the photographer who captured the image. Camera makers should publish full and open descriptions of all parts of the raw image files their camera produce (OpenRAW, 2006, chap. 4)

5. Mackenzie Smith (2005) states:

    Best practice in software development today, especially in areas that are poorly understood like digital archiving and preservation, defines a process by which the system evolves rapidly as our understanding of the problem increases. This is known as "spiral development" (Boehm, 2000), and in practice it means that systems should

be designed with modularity in mind and with the assumption that the code will be all thrown away and recreated often as understanding evolves. Prototypes are created to try new things, and experimentation is encouraged. The assumption is that any attempt to define a "perfect architecture" for the system that solves the entire problem once and for all is naïve and creates too much risk for the organization that depends on it. (p. 10)

6. For the principles guiding the collection of research and heritage materials for the Turnbull Library see the National Library of New Zealand's collection policy (National Library of New Zealand, 2005, Section 10). The Turnbull Library also keeps unpublished materials in traditional and digital formats in its manuscripts and archives, photographs, oral history, drawings, and prints collections.

7. The Turnbull Library's mandate is to build a research collection, focused in particular on New Zealand and Pacific Island studies and rare books. It has the task of comprehensive collecting of published and unpublished material relating to New Zealand and its people (National Library of New Zealand, 2005, Section 3). Government funding is allocated for purchasing material published outside New Zealand.

8. Publishing trends indicate shifts from print to electronic, offline to online, static to dynamic online publishing; the volume of the deep Web is also increasing. Publications take very different forms, and the publishing business models are being transformed and challenged. Ready access to digital collection material is an increasing expectation, and managing the rights of owners appropriately is complex.

9. Observing other collecting institutions' activities, and sharing and validating experience with colleagues in other institutions is crucial, as is cooperation in technological development (see *DPC Forum on Web Archiving* [Digital Preservation Coalition, 2006]). The collaborative work done under the aegis of the IIPC is a good example of this (International Internet Preservation Consortium, 2006).

10. For a discussion of curatorial decision making with regard to selective harvesting see Koerbin (2005).

11. A diagram identifying types of Web sites, their content changes and interactivity can be found in the work of Netarchivet.dk (2003, section 3.1.3).

12. See Crook (2006) for publisher attitudes and speculation on whether this is generalizable to Web material harvested in whole domain without authority; see also European Union recommendations for national strategies and legislation to support preservation of digital cultural heritage (European Union, 2006).

References

Bergman, M. K. (2001). The deep web: Surfacing hidden value. *Journal of Electronic Publishing, 7*(1). Retrieved September 1, 2006, from http://www.press.umich.edu/jep/07-01/bergman.html.

Berkery, Noyes & Co. (2005). *A look at the future of the information and media industries: Trends and opportunities for driving growth.* Retrieved September 1, 2006, from http://www.berkerynoyes.com/PDF/Whitepaper/Aug2005Whitepaper.pdf.

Brown, J. S., & Duguid, P. (2000). *The social life of information.* Boston: Harvard Business School Press.

Cameron, F., & Kenderdine, S. (Eds). (2006). *Theorizing digital cultural heritage: A critical discourse.* Cambridge, MA: MIT Press.

Crook, E. (2006). For the record: Assessing the impact of archiving on the archived. *RLG DigiNews, 10*(4). Retrieved August 31, 2006 from http://www.rlg.org/en/page.php?Page_ID=20962#article0.

Cumings, J., & Mason, I. (2004, September). *The future of the past, the future in the present and beyond.* Paper presented at the Library and Information Association of New Zealand Conference, Auckland, New Zealand. Retrieved September 1, 2006, from http://www.lianza.org.nz/events/conference2004/papers/cumings.pdf.

Digital Preservation Coalition. (2006). *DPC Forum on Web Archiving.* Retrieved September 1, 2006, from http://www.dpconline.org/graphics/events/060612web-archiving.html.

European Union. (2006). *Commission recommendation of 24 August 2006 on the digitisation and online accessibility of cultural material and digital preservation.* Retrieved December

4, 2006, from http://europa.eu.int/information_society/newsroom/cf/document
.cfm?action=display&doc_id=160.

Illien, Gildas. (2006). *Sketching and checking quality for Web archives: A first stage report from BNF* (Bibliothèque Nationale de France unpublished working draft).

International Internet Preservation Consortium. (2006). Software. Retrieved September 1, 2006, from http://netpreserve.org/software/toolkit.php.

Joe, S., & Lala, V. (2006, October). *Web archiving at the National Library of New Zealand.* Paper presented at the Library and Information Association of New Zealand Conference, Wellington, New Zealand. Retrieved September 1, 2006, from http://www.lianza.org .nz/events/conference2006/index.html.

Koerbin, P. (2005, July). *Current issues in web archiving in Australia.* Paper presented at the Open Publish Conference, Sydney, Australia. Retrieved September 1, 2006, from http://www .nla.gov.au/nla/staffpaper/2005/koerbin1.html.

Masanès, J. (2005). Web archiving methods and approaches: A comparative study. *Library Trends, 54,* 72–90.

National Archives of Australia. (1999). *Recordkeeping metadata standard for Commonwealth agencies.* Retrieved August 30, 2006, from http://www.naa.gov.au/recordkeeping/control/ rkms/summary.htm.

National Library of Australia. (2004). *Archiving Web resources: Issues for cultural heritage institutions.* Retrieved September 1, 2006, from http://www.nla.gov.au/webarchiving/program .html.

National Library of New Zealand. (2005). *Collections policy.* Retrieved August 31, 2006, from http://www.natlib.govt.nz/en/about/1keypolcollections.html.

National Library of New Zealand. (2006). *Legal deposit for New Zealand publishers.* Retrieved September 1, 2006, from http://www.natlib.govt.nz/en/services/5legaldeposit.html.

National Library of New Zealand (Te Puna M tauranga o Aotearoa) Act 2003. (2003). Retrieved August 31, 2006, from http://www.natlib.govt.nz/files/Act03-19.pdf.

Netarchivet.dk. (2003). *Experiences and conclusions from a pilot study: Web archiving of the district and county elections 2001.* Retrieved September 1, 2006, from http://netarchive.dk/publika tioner/webark-final-rapport-2003.pdf.

O'Hara, K., Morris, R., Shadbolt, N., Hitch, G. J., Hall, W., & Beagrie, N. (2006). Memories for life: A review of the science and technology. *Journal of the Royal Society Interface, 3,* 351–365.

OpenRAW. (2006). *The 2006 OpenRAW survey: A report on the experiences, requirements, beliefs, and preferences of photographers and imaging professionals regarding RAW imaging technology.* Retrieved October 28, 2006, from http://openraw.org/survey.

Phillips, M. (2005). What should we preserve? The question for heritage libraries in a digital world. *Library Trends, 54,* 57–71.

PREMIS Working Group. (2005). *Data dictionary for preservation metadata: Final report.* Retrieved December 12, 2006, from http://www.oclc.org/research/projects/pmwg/premis-final .pdf.

Rabinovitz, L., & Geil, A. (Eds). (2004). *Memory bytes: History, technology, and digital culture.* Durham, NC: Duke University Press.

Smith, M. (2005). Exploring variety in digital collections and the implications for digital preservation. *Library Trends, 54,* 6–15.

State Library of Victoria. (2005). *Digital preservation policy.* Retrieved September 4, 2006, from http://www.slv.vic.gov.au/about/information/policies/digitalpreservation.html.

Woodyard-Robinson, D. (2006). *Decision tree for selection of digital materials for long-term retention.* Retrieved September 1, 2006, from http://www.dpconline.org/graphics/handbook/ dec-tree.html.

Ingrid Mason is Digital Research Repository Coordinator, New Zealand Electronic Text Centre, Victoria University of Wellington. Until recently she worked at the Alexander Turnbull Library and the Innovation Centre, National Library of New Zealand, on developing the library's infrastructure and practices for collecting, preserving, and providing access to digital heritage materials, in particular to electronic publications. She has also been a reference librarian and intranet coordinator at the

Powerhouse Museum (Sydney, Australia), a lecturer in the Masters of Library and Information Studies program at Victoria University of Wellington, and a business analyst, mostly in the cultural heritage sector. Among her recent publications are a joint paper for the 2006 conference of the Library and Information Association of New Zealand Aotearoa outlining the development of the Web Curator Tool, an open source tool to manage selective Web harvesting, and a chapter, "Cultural information standards: political territory and rich rewards" in *Theorizing Digital Cultural Heritage: A Critical Discourse.*