Digital Libraries: Preserving Information Resources for Contemporary South African Culture and Scholarship

Dale Peters

Introduction

The digitisation of our cultural heritage brings together various sectors of the academic community in a manner unprecedented in traditional library structures. Scholars are creating or using electronic resources to further their research; distance learning models prompt teachers to gather Web resources in an online learning environment and publishers are integrating print with digital editions to reach wider audiences. The support of computer and information specialists in the application of new technologies to develop and manage Humanities resources is sustained by librarians and archivists seeking to improve access to and the preservation of digital information that represents resources of contemporary culture and scholarship.

This dynamic collaboration has not been without a painful birthing process in the transformation of higher education. Although not limited to this country, the effects of public policy on education are compounded in the South African effort to redress past imbalances. The near-death experiences of restructuring in the Humanities has forced us to rethink the relationship of our work to society, and to re-establish our prestige in higher education by demonstrating a wider contribution to society. With the global trend towards a knowledge-based economy, the future wealth of nations is calculated increasingly on the education of its people, and not on its natural resources (Hughes 2000). To meet the demand for knowledge, seen as a key to prosperity, the goal of higher education is to provide opportunities for lifelong learning, to create second chances. To achieve the goal of this social contract, the collaborative effort of all those involved in Humanities research is aimed

Alternation 9,2 (2002) 109 - 118 ISSN 1023-1757

at high quality education and information access to all people, regardless of social class, the ability to pay, or their location. In this paper I want to reflect on the role of digital libraries in South Africa in the creation and maintaining of digital information resources and cultural heritage for academic scholarship in the Humanities.

Creating Resources of Digital Information

Preservation is amongst the core activities of the Campbell Collections in response to the deleterious effects of the humid climate in the subtropical southeast African coastal environment. The functions of preservation management encompass areas of specialised paper conservation, collections maintenance, environmental control, disaster response, and in reformatting, previously limited to photocopies and microfilm. Recent trends in the field of preservation have seen the development of digital technology as a preservation management strategy aimed at the reduced handling of rare objects, in a manner ideally suited to the ultimate aim of libraries and archives—in providing enhanced research access.

For the past five years, the Campbell Collections has been actively involved in digital library research and development activities in several areas, including digital imaging, metadata creation, Web publishing and digital preservation. The digital library initiative was developed to utilise digital methods to preserve and to make accessible fragile and under-utilised visual resources; to promote the use of digital media in the university and wider scholarly community; and to conduct research that advances the creation, management, preservation and use of digital resources.

Digital technology lends itself to the objectives of the library community, with a long history of electronic co-operation, in joint catalogues, and levels of interoperability enabling the interrogation of remote databases. The advantages of digital technologies are apparent in providing remote access to a unique collection of primary research materials. Traditionally comprising books, documents, artefacts and works of art of significant research merit, the Campbell Collections attracts international researchers, students and scholars of history, sociology and of the wider cultural heritage of KwaZulu Natal.

The digital conversion of the historic photograph collection is a case in point. A resource used heavily for research purposes, repeated browsing of the photograph albums has resulted in an alarming deterioration in their

physical condition. While the primary objective of limited handling is achieved in digital conversion, the concept of digital libraries is far greater than reformatting and digital preservation is achieved in complementary strategies. Firstly, digital conversion to international benchmark standards enables fast download speeds for satisfactory on-line access; while large uncompressed masters are committed to archival storage. 'Metadata' are created for each image, based on the Dublin Core Metadata Initiative[1] for electronic resources. 'Metadata'—the data about the data—both descriptive and technical, play a significant role in digital preservation. In conjunction with emerging standards (such as Dublin Core), the creation of metadata is amongst the most promising ways to keep digital resources 'alive' over the years and decades, by recording file formats and data types that typify the rapidly evolving landscape of software and hardware platforms in the transient space of information technology. Secondly, metadata are useful for all kinds of resource discovery, in using and accessing digital collections in a user-friendly way. Other emerging standards such as XML (Extensible Markup Language)[2], TEI (Text Encoding Initiative)[3] and EAD (Encoded Archival Description)[4] have found similar application in the digital library. Finally, the conservation treatment of individual albums and the construction of protective phase box enclosures is conducted, with due consideration for the specific requirements of environmental control for the archival storage of artifacts comprising photographic materials. The intellectual content historic photograph collection can be viewed at http://khozi2.nu.ac.za/digimage.html.

Valuable experience has been accumulated at the Campbell Collections in the creation, management, preservation and use of digital resources in the role of host to the South African national co-operative project, DISA. This pilot project, funded by an Andrew W. Mellon Foundation grant,

[1] Dublin Core Metadata Initiative (http://dublincore.org/about/).

[2] XML is a subset of SGML, the Standard Generalized Markup Language (ISO 8879) (http://www.gca.org/whats_xml/default.htm).

[3] Text Encoding Initiative, an international project to develop guidelines for the encoding of textual material in electronic form for research purposes (http://www.tei-c.org/).

[4] Encoded Archival Description is a nonproprietary encoding standard for machine-readable finding aids such as inventories, registers, indexes, and other documents created by archives, libraries, museums (http://lcweb.loc.gov/ead/).

aims to gather from research institutions around the country, to digitise and to present a representative selection of texts, entitled South Africa's Struggle for Democracy: Anti-Apartheid Periodicals, 1960-1990.

Periodical titles were selected by a national committee comprising academic scholars, librarians and archivists, to represent the wide spectrum of political views published during these years, and a range of subjects such as trade unions, religion, health, culture, and gender. Publications reflect both black and white viewpoints, and an attempt has been made to represent distinctive regional variations. Some of the publications were, by nature of frequent banning orders, often short-lived and thus by necessity, of limited distribution. These factors lend a certain rarity value in that the publications are generally not well represented in research collections. The digital imaging project will of course, also preserve the original copies through a reduction in the handling of fragile objects[5] (DISA 1998).

Making digital copies of these periodicals will facilitate access to an important resource on this period of South African history. Original copies are gathered together from collections around the country, indexed to article level and linked to the full text and images of the periodicals. A typical request from the Africa Collection, Hoover Library at Stanford University bears witness to global access to the broadest possible community of users via the Internet. The page images are presented on a dedicated website (see http://disa.nu.ac.za).

Preserving Intellectual Content

Preservation digitisation has focussed on reformatting rapidly deteriorating materials, with the aim of preserving the intellectual content, as separate from the medium or carrier. Because one is able to make repeated copies of digital files without the informational loss associated with analogue copies, the impact of the technology is beguiling. No longer does the life expectancy of materials seem limited to the durability of the medium, since the digital data can be copied repeatedly without loss.

The separation of the information and the object however, takes on a new meaning in the digital environment. Complete digital conversion comprises two process, the capture by scanning of digital data, and the

5 DISA 1998. South African Digital Imaging Project: A Proposal to the Andrew W. Mellon Foundation.

creation of the associated metadata, wrapped as a digital object and stored using containers to hold the objects and related software modules. The integrity of the digital file is threatened by alterations to the contextual data file, and the loss of the linking mechanism between the two is equivalent to the deletion of the contextual information (Beamsley 1999).

Methodologies for Digital Preservation

Digital resources are tomorrow's heritage. The second stage of this discussion is the consideration of the role of digital libraries to develop strategies and methodologies for digital preservation to ensure long-term access to that heritage. Preservation strategies can be described within a framework for the long-term storage, maintenance and accessibility of a digital object.

Notable differences between digital and paper-based materials present new challenges in maintaining access to digital resources. Digital resources are machine-dependent. They require specific hardware and software for retrieval. The rate of change in technology demands that actions are taken within a timeframe much shorter than that for traditional materials. Strategies for long-term preservation are integral to the process of creation as noted above. Although digital materials are threatened primarily by technological obsolescence, physical deterioration of the unstable media is inevitable without suitable storage conditions and deliberate preservation management (Jones & Beagrie 2000).

Three potential strategies defined for digital preservation are technology preservation, technology emulation and the migration of digital information (Hendley 1998). Technology preservation involves data storage on a stable medium, and in a suitable environment; the refreshing or copying of the data to new media as required; maintaining the data integrity of the containerised digital data and associated contextual information file; preserving the original software application/s required to access the data types preserved; providing the operating systems on which the original applications ran and preserving the hardware platform on which the operating system was designed. The predictable result of technology preservation is a veritable museum of obsolete computing equipment, and is considered only as a relatively desperate measure, where digital resources cannot be converted to independent formats and migrated forward.

Technology emulation differs from the technology preservation strategy in the design of emulator platforms that mimic the operating system

and hardware environment in which the original applications were designed (Rothenberg 1999). A highly specialised strategy, the success of emulation is associated with a high level of risk in the dependence on the technical ability a software engineer, if indeed such a person could be found, considering the unlikely commercial viability of emulating a specific environment.

The third strategy of digital preservation, the migration of digital information, is based on the premise that it is only worth preserving those resources that can be accessed on current computer systems. This is achieved by ensuring that data is created or migrated to a format that is platform independent and then migrated from one generation of computer technology to the next. Given the rate of change in that landscape, digital migration is not an option—it is an essential strategy comprising a number of tactics.

The transference of data from less stable to more stable media is an irony in itself. The market forces driving changing modern media can be seen in an inverted relationship of a decreasing life expectancy to an increasing information density (Conway 1999). A comparison of the Sumerian clay tablet to the magnetic disks of mainframe computers leaves one pondering the perspectives of progress. Indeed, paper and microfilm probably constitute the more stable option for media change, and the process is cheap and easy to implement. The danger lies in 'flattening', or the loss of data structures, graphics capabilities, and indexing relationships integral to complex data systems and which serve in the authentication of the document. Changing media is best prioritised as a backup strategy for digital preservation management.

Common experience of word processing packages has indicated that a reliance on backwards compatibility is not always sufficient, and interoperability is more readily achieved in the use of a common interchange format, the American Standard Code for Information Interchange (ASCII) or Rich Text Format (RTF). The loss of formatting is noticeable in figures and footnotes, which, from the author's viewpoint, probably constitutes the loss of valuable content. Always concerned with risk management, a preservation management strategy reliant upon interoperability and backwards compatibility would be dependent on the unknown future of the interchange formats selected, that may cease to be supported or replaced by new, richer formats.

The final migration strategy, fortunately also recommended as most appropriate for digital libraries, is the migration to standard formats that encode the complexity of the original formats. Text documents would thus

conform to standards like SGML (ISO 8879) and image files conform to tagged file formats (TIFF) and standard compression algorithms (JPEG). Under this strategy, the onus is on the digital project manager to define preferred formats at the time of the creation of digital resources, and to limit the number of variable formats that may become necessary to migrate to future generation of computer technology. Other key decisions need to be made as to whether the resources is to be stored as raw data, in which the representation of the information would be secondary to the presentation as a table or graph; or whether, as in the case of digital libraries, the objective is to preserve as much of the original format and presentation as possible to ensure its archival integrity.

New Avenues of Collaboration

The collaboration of a wide community of information professionals has been largely responsible for developing new strategies for preservation management of the digital information resources developed at the Campbell Collections in support of the core function of academic research. This is a natural extension of the role of the Campbell Collections in organising and archiving knowledge systems in the range of fields in the Humanities and has served to develop a gravity centre for the building of the technical capacity of staff.

Some of the factors noted impeding such collaboration merit consideration, and include problems of interaction within the organisation related to territoriality. A lack of informed support has been identified in the radical transformation of the information environment. While the digital revolution is unlikely to change the art, science and enduring values of librarianship, it has separated information from the physical locality of the library and changed the way we view academic libraries especially, as places to study, to get on-line and to pursue the normal rites of male-female bonding among students. The generational fear of change associated with 'technophobia' has steadily decreased, possibly by natural attrition, possibly as a result of its positive identification as the psychoneurotic barrier to technological innovation in cultural institutions (Greenberg 1997).

The digital library provides the infrastructure that brings together scholars, academic researchers, publishers, librarians and archivists to foster the exchange of ideas and technical skills to extend the use of digital resources as techniques and tools in Humanities research. Further collaboration is envisaged to support those members of the academic community who have

already made a commitment to academic computing. The development of future digital projects will seek joint funding for a diverse, interdisciplinary set of humanities research fellows, working in a broad range of relevant fields, including archaeology, art, art history, history, information studies, languages and linguistics, literary studies, music, the performing arts, as well as work detailing techniques and issues associated with the creation and use of digital texts, databases, images, sound, video and digital mapping. The digital library could support academic research projects in providing access to high quality scanning equipment, computer workstations, multimedia development software, and personal technical support. Projects carried out may include the creation of image databases, development of multimedia research and instructional materials and the digitisation of various media.

Digital preservation policies and practices are not currently well developed in South African libraries, and few have assumed responsibility for preserving materials in a digital form at all. Yet, a survey conducted in 1998 by the Research Libraries Group noted that most institutions which began acquiring or creating digital data in the previous five years found that they could not access all of their materials because they lacked the operational expertise or technical capacity to access some storage media in their holdings (Hedstrom & Montgomery 1998).

Digital technology is a central leadership issue of the current debate on information management. There are those who think that leadership on technological issues is simply a matter of establishing procedural guidelines. Others feel that the rapid rate of change and the sheer technological complexity of the digital environment render librarians and archivists helpless in influencing technological development. Both perspectives are misleading. The former ignore the market forces in rapid IT development, while those who prefer to 'wait and see' how digital technology develops shirk their responsibility to contribute to the debate.

Preservation in the area of digital technology is a shared responsibility. In digital libraries we are working with people involved in the Internet, the Web, multimedia. document imaging and other technologies that serve a constituency far wider than the University library system, or a Humanities research project. Each has a different concept of digital preservation, but it is in the digital library that their various efforts become focused on long-term access. Digital libraries naturally draw together the composite skills of libraries, museums, archives and galleries. Publishers are major players in the digital arena, ranging from scholarly to large commercial

publishers. Then there are hardware and software companies vying for their market share; computer scientists from corporations, academic departments and government associations. Academic staff are represented in the digital library arena in almost all disciplines, in fields nominally unrelated to computer science, but involved through a need to use digital technologies as a tool towards their own ends. All of these groups are involved in a dynamic field of digital library research, development and applications, and the diversity presents a breadth of knowledge for an exciting synergy in the creation of innovative resources of contemporary culture and research.

There is an ongoing need to work together as a community to establish measures to build capacity and increase expertise with issues of digital technology in higher education. A strong voice is needed to express the need for high quality education and information access to all people. A collaborative effort will most efficiently drive the development of needed products, to grow a demand for commercial services and to build technically sustainable solutions that support international standards to meet the needs of digital preservation that will effectively secure long-term access to digital information resources

Campbell Collections
University of Natal

References

Hughes, CA 2000. Information Services for Higher Education. *D-Lib Magazine* 6,12: 9. http://www.dlib.org/dlib/december00/ hughes/12hughes.html.

Dublin Core Metadata Initiative. http://dublincore.org/about/.

DISA 1998. South African Digital Imaging Project: A Proposal to the Andrew W. Mellon Foundation.

Beamsley, TG 1999. Securing Digital Image Assets in Museums and Libraries: A Risk Management Approach: In Sandore, B (ed): *Progress in Visual Information Access and Retrieval.* Champaign, IL: University of Illinois Graduate School of Library and Information Science.

Jones, M & N Beagrie 2000. Preservation Management of Digital Materials: Workbook. A Pre-publication Draft. http://ahds.ac.uk/ manage/licstudy.html.

Hendley, T 1998. Comparison of Methods and Costs of Digital Preservation British Library Research and Innovation Centre Report No.106. http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html.

Rothenberg, J 1999. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation.* Washington DC: Council on Library and Information Resources.

Conway, P 1999. Preservation in the Digital World. In School for Scanning. Chicago: Northeast Document Conservation Centre.

Greenberg, D 1997. Technophobia, Papyrophilia and the Real Thing: Psychoneurotic Barriers to Technological Innovation in Cultural Institutions. *Association of Computing Machinery Newsletter* Feb 1997.

Hedstrom, M & S Montgomery 1998. *Digital Preservation Needs and Requirements in RLG Member Institutions.* http://www.thames.rlg.org/preserv/digpres.html.