Will your research assets survive the ravages of time, redundancy, mismanagement and decay? Planning for digital preservation within higher education institutions

Ina Smith
University of Stellenbosch
ismith@sun.ac.za

Abstract: *This paper presents a content analysis of a selection of existing literature about digital preservation policies. The researcher examined the digital preservation policies that are currently being implemented by four representative institutions on four different continents and arrived at specific conclusions. The institutions reviewed in this paper are Hong Kong University of Science and Technology (Asia); University of Texas Library (United States of America); National Library of Australia (Australia); UK Data Archive, University of Essex (United Kingdom). Although the survey was undertaken in April and May 2010, the primary purpose of the content analysis was to provide a rational basis for a sound digital preservation policy for the institutional repository of the University of Stellenbosch in South Africa. The paper defines digital preservation, examines the rationale behind digital preservation policies, elucidates issues that need to be resolved before long-term access to digital research objects within higher education institutions can be guaranteed. It also describes all the essential elements that need to be included in a digital preservation policy. The researcher illuminated the digital preservation policies of the above-mentioned institutions by comparing them with the requirements that have been documented by the PARS working group. The paper concludes with recommendations about which elements should receive particular attention when a digital preservation policy is constructed for the use of institutional digital repositories.*

Keywords: *digital preservation policies, sustainability, policy formulation, long-term preservation, digital objects, research asset management, institutional repositories*

1. Introduction

Since no one can predict future developments with absolute accuracy, we cannot be sure that the way in which we currently preserve our digital research output will enable us, in future years, to access the digital files that have been accumulating in our repositories. This very uncertainty compels us to take whatever measures we can before the situation deteriorates beyond the point of no return. In a podcast recorded by the Repositories Support Project (2009), Kevin Ashley of the University of London Computer Centre said: "Repositories need to think about preservation

even if in doing this they decide it's not important." But, in the opinion of McGovern (2009), merely *thinking* about this inevitable problem is insufficient. McGovern writes: “A high-level policy that defines an organization’s commitment to digital preservation is an *essential* component of an effective and sustainable program.” If your institution has already committed itself to preserving research in a digital format, a detailed policy that describes the ways and means that it proposes to accomplish this goal, is an essential and non-negotiable precursor to ensuring the successful long-term preservation of digital records.

Anderson (2005) and Strodl et al. (2007) are also of the opinion that preservation planning needs to be undertaken within each organization that works with digital objects. PLANETS (the acronym for the EU project for “Preservation and Long-Term Access via Networked Services”) has identified planning as an indispensable part of the digital preservation process as a whole in the context of modern archival standards. It also constitutes a core functional entity in the OAIS model, the ISO-adopted Reference Model for an Open Archival Information System, which is a common reference model for all archival activities (Strodl et al., 2009; Woollard, 2009).

In their paper entitled *Digital preservation: are repositories doing enough for preservation*?, the authors of the *Repositories Support Project (2009)* issued the following challenges: “It's not enough for repositories to sit back and wait for preservation services – they have to specify what they want, to be user-centred, to establish policies." This is an enormous challenge, the difficulties of which are described in more detail by Lam and Chan (2007, p. 322) when they write: “Installing institutional repository software such as DSpace™ is straightforward, but tailoring software and setting up policies and procedures to make it work effectively in one’s institutional environment are uphill tasks.”

What all this emphasizes is that we need to design a system that takes us *beyond* the limitations of institutional repository software so that we will not have to rely on it completely to guarantee our future access to valuable digital objects. Rice University bases its approach to digital preservation on Dspace™’s commitment to digital preservation (Wise 2007). And while making use of open source technology is one of the essential features of a successful digital archive, we still need to take into account what may happen beyond the artificial boundaries imposed by our current systems.

Since digital objects are far less tangible and potentially more ephemeral and destructible than printed objects, they are also far more difficult to preserve and secure with confidence. This in itself creates many challenges that should not be ignored. All electronic content has to be monitored and managed in such a way that its accessibility and usability can be assured far into the future (Strodl et al., 2007). This can only happen if a dedicated policy and plan of action are currently in place in an institution that values the integrity of its digital records.

Policy not only helps to establish trust and confidence, it also helps to minimise or entirely remove risks from the entire system (risk management). Because a sound policy offers consistent guidelines with regard to the procedures and protocols that need to be followed by those responsible for preserving digital records, such a policy also ensures the making of consistently correct decisions. According to the PARS working group (2007): “Digital preservation policies document an organization’s commitment to preserve digital content for future use; specify file formats to be preserved and the level of preservation to be provided; and ensure compliance with standards and best practices for responsible stewardship of digital information.” But such a digital preservation policy cannot exist in isolation. It will only be truly effective if it is consistent with the needs and purposes of core institutional business drivers and strategies.

2. Problem statement and research methodology

This research report identifies various key elements that are either included in or excluded from existing digital preservation policies in the following four institutions:

- Hong Kong University of Science and Technology (Asia)
- University of Texas Library (United States of America)
- National Library of Australia (Australia)
- UK Data Archive, University of Essex (United Kingdom)

The researcher has conducted a content analysis of the preservation policies of these four institutions, and has, in the process, identified similarities and differences. She has also collected additional useful data by means of e-mail correspondence with staff who were responsible for the implementation of the preservation policies of these institutions.

The researcher’s prior assumption was that this analysis would be able to provide a basis for the development of coherent and practical digital preservation policies in local South African and African universities and institutions that need to preserve their digital records. By identifying the key elements in the preservation policies of the institutions selected for review, the researcher emphasised those elements that cannot be ignored by local professionals who are responsible for compiling effective digital preservation policies for their own institutions. At present, there is no South African higher education institution that has adopted a policy for the preservation of digital research output that, if applied, will preserve and secure research records in digital format as a permanent part of the institution’s repository. This is indeed an unfortunate situation because digital assets that were once preserved by means of older forms of computerised technology (such as research texts that were created in WordPerfect and then stored on floppy disks) cannot currently be retrieved except by special efforts that require a considerable investment of time,

money and ingenuity. Another example of inaccessibility occurs when our own researchers publish research that is based on databases that are inaccessible to us.

Even though this study by no means represents a complete description of the techniques and methods of those who preserve digital records in institutions throughout the world, it nevertheless presents a basic analysis of the preservation policies of the four reputable institutions mentioned above – as these are perceived and interpreted by the author.

3. Contextualising digital preservation in institutional repositories

The literature indicates that there are various degrees of uncertainty, lack of experience and an inability to reach consensus about how to proceed with digital preservation processes within institutional repositories (Jubb, 2007; National Library of Australia, 2008). The main emphasis has been placed on open access policies and the handling and dissemination of research materials. This is evident in, for example, Cranfield's approach, about which Bevan (2007) notes: "QUEPrints is more concerned with access than preservation" and "The current answer is to ensure that all material is in PDF format."

Many practitioners regard the keeping of duplicate copies of a file (exemplified in the LOCKSS principle, namely, "Lots of Copies Keep Stuff Safe") as a sufficient and adequate means for guaranteeing future access to any file. But this practice does not take into account the fact that certain file formats rapidly become unreadable or obsolete as computer technology and development expands exponentially over time. Greig and Nixon (2007) also refer to PDF redirected content. When a paper is not held locally, the only connection to it is represented by a link to an external repository (and the reason as to why the paper was not deposited locally is recorded). Needless to say, such a procedure does not guarantee the preservation and integrity of a full text data file over a long period of time. In fact, it cannot be regarded as a digital preservation strategy at all.

Measures for the long-term access to and preservation of digital archives is also not regarded as a prime funding priority of specifically South African universities and institutions. This is exemplified by the fact that while the National Research Foundation sponsors researchers and offers grants for local research, it at no stage requires a researcher to make provision for long-term access to his or her research products.

In the following section, the researcher examines the concept of digital preservation in terms of how it has been defined in the literature, and then offers a critical analysis of existing digital preservation policies.

3.1 The definition of digital preservation

Jantz and Giarlo (2005) define *digital preservation* as "the managed activities necessary for the long term maintenance of a byte stream (including metadata) sufficient to reproduce a suitable facsimile of the original document and for the continued accessibility of the document contents through time and changing technology". The digital preservation policy of the National Library of Australia (2008) notes there are a variety of challenges that have to be met if digital information resources are to remain accessible and in good condition over time.

The members of the PARS (Preservation and Reformatting Section) working group which met during the American Library Association's 2007 Midwinter Meeting prepared themselves by studying a number of resources in order to familiarize themselves with the most crucial elements of digital preservation that had already been identified by a broad spectrum of individuals and agencies. After their deliberations, they produced three definitions that dealt with short-, medium- and long-term preservation respectively. The following definition refers to medium-term preservation:

> Digital preservation combines policies, strategies and actions to ensure the accurate rendering of authenticated content over time, regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time (ALA 2007).

Moore (2008) conceptualises *preservation* as a method of communicating with the future. According to Moore, preservation implies that information that is readily comprehensible and accessible today has been prepared and formatted so that it can be transmitted to an unknown system that will become operative in the future and that will enable it to be accurately interpreted and displayed for future users. It may be taken for granted that future systems will not only make use of different kinds of hardware and software, and that they will also utilise different standards for encoding information. In order to be effective, we therefore need to prepare digital texts for future use by not only providing a description of the format of the relevant information (byte stream or full text files), but also by providing a detailed description of the environment and technology that is presently in use to manage and read those files.

But exactly how do institutions plan to “communicate with the future” when it comes to their digital assets?

3.2 Digital preservation strategies

If we are to ensure that digital materials will survive for the next one hundred years, for example, we need to clarify the methods and procedures that we use to manage these materials today. Most of these steps involve storing digital objects with open systems and applying open standards and open digital formats (Gibson, 2010).

The success of the Internet can be attributed to the open standards that are used for data management and storage. The universal success of this system means that we can confidently adopt it as the best possible exemplar as we seek to choose the most effective methods for digital preservation. There are already a few organisations that are investigating the use of open standards for computing services. One of these is the Open Source Software Movement (http://www.opensource.org). These standards need to be stored locally so that they can be easily referenced by future users and digital archivists.

The success of the PDF format for the storage of digital documentation also suggests another method for achieving digital preservation. The Unified Digital Format Registry (UDFR) (http://www.udfr.org) is currently identifying digital formats and creating a registry in which these formats will be preserved. Gibson (2010) suggests that the use of open formats should be actively encouraged and supported so that all data will ultimately be converted to open formats. The preservation strategy and policy of the University of Essex, for example, clearly states that its preservation policy and strategy is squarely based on open and available file formats.

The use of open systems is the most logical method of guaranteeing future access to digital objects. The more complex a system is, the more difficult will it be to maintain its integrity and accessibility in the remote future. In line with these practices, the researcher also recommends that the customization of existing open source systems should be kept to the minimum, and that all changes to the basic open source format should be officially documented so that future upgrades and migrations will be easier to effect. As a precautionary measure to forestall whatever problems might arise in the remote future, many Information Technology departments are currently purchasing systems from large proprietary firms. Since these suppliers will in all likelihood not even exist in a hundred years from now, the data that has been generated by them will be lost unless the hardware that displays the data has been preserved in working condition (Gibson, 2010). Gibson (2010) also suggests that – wherever possible – open computer systems that are characterised by the least possible degree of complexity should be used. Other important cautions are that as little reliance as possible should be placed on external vendors for support

and maintenance, and that all systems that are purchased should be based on the OAIS (Open Archival Information Systems) Reference Model.

The PARS working group (2007) also identified various other elements connected with content creation, integrity and maintenance that need to be taken into account as part of a rational and effective digital preservation policy for an organization. The author of this paper has freely used and interpreted these elements in her analysis of the digital preservation policies of the institutions mentioned in the introductory section of this paper. In those cases where the institution concerned offered no solution for an identified problem, or where no information was available at the time of the compilation of this paper, such lacunae are simply indicated by a blank space in Table 1 (Addendum A). This does not necessarily indicate that the problem concerned is not being addressed by the particular institution, or that it does not plan to address these problems in the future.

4. An analysis of digital preservation policies

In this paper, the author analyses and compares the digital preservation policies of the Hong Kong University of Science and Technology (Asia), the University of Texas Library (United States of America), the National Library of Australia (Australia), the UK Data Archive at the University of Essex (United Kingdom) in terms of the framework compiled by PARS (Preservation and Reformatting Section) (2007).

The researcher noted that, in most instances, each of these institutions had a digital preservation policy that was separate and distinct from the overall institutional repository policy of the institution, except for the Hong Kong University of Science and Technology Library, whose digital preservation policy formed a part of the overall institutional repository policy of the university. The UK Data Archive digital preservation policy is based on the principles of open and available file formats, data migration and media refreshment. While this archive also addresses the issue of sustainability, it is committed, in addition, to funding all operations that are engaged in preservation, management, technical infrastructure, financial planning and staffing infrastructure. This archive has established a “rolling” planning scheme which takes into account the probable lifespan and availability of computer equipment and storage media so that it will be in a position to obtain the necessary upgrades (Woollard, 2009). In comparison to all the other institutions selected for this study, the policy of the UK Data Archive was by far the most comprehensive.

It is the view of the PARS working group (2007) digital preservation strategies and actions should be concerned with content creation, integrity and maintenance. Table 1 (Addendum 1) provides a comparative summary of the digital preservation policies in terms of the PARS guidelines.

4.1 Content creation

The technical specifications provided by the UK Data Archive were both clear and comprehensive, although they made no reference to recommended file formats. The University of Texas Library, on the other hand, provided a specific list of file formats that they preferred, even though they have committed themselves to making every effort to preserving an inventory of all file formats. Their policy is to create a derivative file in every case where non-preferred formats have been used. In such cases, both the original as well as the derivative file are preserved for posterity. The Hong Kong University of Science and Technology (hereinafter referred to as HKUST) simply stated that it is their policy to retain all items indefinitely. They also state that even though they try to create the conditions that will ensure readability and accessibility in the remote future, they cannot provide any guarantees when it comes to accommodating unusual file formats. While the PDF file format has been adopted by HKUST, the National Library of Australia supports all international standards.

All institutions selected for the study indicated that they were committed to the preservation of the byte stream that was originally submitted to them.

The University of Texas Library, the National Library of Australia and the UK Data Archive indicated that they collate and preserve whatever descriptive, administrative and structural metadata is necessary to ensure future access. The UK Data Archive policy contains detailed information about how quality control is applied during all processes (they base their procedures on the OAIS Reference Model) (Woollard, 2009). It is interesting to note that all sections and staff of the UK Data Archive have a role to play in the implementation of the archive's preservation policy, and that the duties and obligations of all the members of staff concerned are clearly defined. The UK Data Archive has activated a pre-ingest function to ensure the quality, comprehensibility and accessibility of all information packages. This method reduces the running costs of the ingest process because it is performed strictly according to predetermined standards with regard to, for example, content, confidentiality, ethics, legal issues, data formats (Woollard, 2009). During the ingest function (which is based on the OAIS Reference Model), all the elements of the deposited files are transformed into a valid preservation format for the specified data type. The files that are intended for preservation are then copied onto a different machine so that the ingest and preservation directory structures can be created (Woollard, 2009).

4.2 Content integrity

The OAIS (Open Archival Information System) Reference Model is an international standard (ISO 14721:2003) and a conceptual framework which describes the components necessary for a digital archive. All four institutions have indicated that their policies comply with the requirements of the OAIS Reference Model. What makes the policy of the UK Data Archive unique is that the most recent version of their policy (2009) is the first to explicitly use OAIS terminology (Woollard, 2009). None of the other institutions have as yet integrated OAIS terminology into their policies.

It was noted earlier in this paper that a digital preservation policy should not exist in isolation – a proposition that has already been acknowledged by both the National Library of Australia and the UK Data Archive. Both these institutions indicated that their digital preservation policies should be read in conjunction with their other library policies, strategic documents, acts, laws and regulations, as well as current best practice. Since a glossary is useful and sometimes indispensable for explaining the meaning of technical jargon, the University of Texas Library and the UK Data Archive have included such a glossary among their documents.

The management of the data collections accessioned by the UK Data Archive are regulated by complex legal and regulatory frameworks (Woollard, 2009). Submitters are issued with a legal license during the ingest process, and the UK Data Archive does not ingest materials whose ownership is unclear or materials about which unresolved rights issues exist.

Both the National Library of Australia and the UK Data Archive acknowledge that since they cannot function in isolation, they are keen to cooperate with other institutions that have to cope with similar problems as they undertake the preservation of their digital assets.

Enduring or persistent identifiers are a requirement for guaranteeing future access to items in a digital repository. For this purpose, the University of Texas Library and HKUST make use of the CNRI Handle System ® (see http://www.handle.net). Even when items have been withdrawn from the repository, URLs will continue to point to “tombstone” citations so that links will remain and unbroken and item histories will be retained. These identifiers are therefore resolvable in perpetuity, and will remain valid even once the content concerned has migrated to a new system. While the handle system currently in use by the National Library of Australia remained unclear to the author, their policy statements confirm that they do assign enduring identifiers to all items.

The provenance and history of changes effected in all objects are recorded in the metadata. This procedure is applied by the University of Texas Library, HKUST and the UK Data Archive. While the UK Data Archive (Woollard, 2009) places a very strong emphasis on authenticity, integrity and reliability, it does not ignore the factor of usability. The UK Data Archive in fact resorts to “soft” deletion as default method (i.e. while certain references to the withdrawn content are

deleted, the content itself is not). They also update the administrative metadata as well as the external view of the catalogue record to reflect any change in the status of the item. They will therefore provide whatever information might be needed for a potential user to see why the item was originally withdrawn, the dates of its availability, and (where appropriate) the reasons why it was withdrawn (Woollard, 2009).

The UK Data Archive has also instituted verification mechanisms to assure the integrity of the content that is submitted to them.

Both the National Library of Australia and the UK Data Archive have put various security measures in place to monitor items in the repository. The National Library of Australia is continuously alert to the possibility of significant threats and malware that would threaten the integrity of their collections. They therefore store and manage their digital collections by means of procedures that ensure their integrity. These procedures include adequate backups and a spectrum of disaster recovery safeguards. In order to comply with national standards for the management of information security, the UKDA follows BS ISO/IEC 27001: 2005; BS ISO/IEC 27002: 2005, BS 7799-1: 2005, Cross Government Actions: Mandatory Minimum Measures. These provide secure networking and communications equipment for the provision of adequate connectivity, and restrict users to valid Mac addresses. They also have a facility for segmenting the network for switched separated firewall connectivity.

Audits and audit trails are invaluable for providing information about the authenticity and integrity of items within a collection. Routine audits are conducted by the University of Texas Library, the National Library of Australia and the UK Data Archive. As far as the UK Data Archive is concerned, the ingest process includes an unbroken trail of actions that guarantees the authenticity and integrity of any data collection. They have also taken account of the necessity for depositor accountability – as may be seen from the fact that they inform all depositors about all actions that have been undertaken within the UK Data Archive before the data collection is released to any wider community. This policy is carefully monitored by regular planned audits that assess the extent to which the requirements of the policy are being implemented. Such monitoring includes a biennial benchmarking of the UKDA preservation agenda. They also preserve an audit trail of all alterations that have been effected to the preserved and disseminated versions, and the connection of these changes to the original deposited version are clearly indicated (Woollard, 2009).

4.3 Content maintenance

The extent to which the various institutions reviewed in this paper possessed a robust computing and networking infrastructure, was not clear to the researcher. HKUST (Lam & Chan, 2007), for

example, run their repository on a single database, and they haven't adopted the multiple databases model set in place by the California Institute of Technology. A section entitled “About the IR” notes that a robust, standard, compliant infrastructure is used to run the repository.

Both the University of Texas Library and HKUST use DSpace™ open source software, which is OAI-PMH (Open Archives Initiative-Protocol for Metadata Harvesting) compliant. This makes their databases inter-operatible so that metadata harvesters and other search engines can retrieve metadata from these repositories. OAI-PMH is a harvesting protocol for sharing data between online services such as repositories (See http://www.openarchives.org/pmh).

Chinese characters need to be accurately recognised and correctly preserved. In order to accommodate special characters, all institutions use the Unicode character encoding scheme. Because it is the most comprehensive encoding scheme available, it is also by default part of any DSpace™ installation unless it is deliberately changed or manipulated during the installation process (Lam & Chan, 2007).

The policy of the UK Data Archive requires them regularly to monitor and update their hardware and software so that long-term access can be assured. Back-ups are also performed by all four institutions on a regular basis. At HKUST, for example, such backups serve to establish the provenance and priority of ideas and intellectual property since all work is registered with an allocated date stamp.

The UK Data Archive preserves five versions of the complete system: a main near-line copy (the main preservation server), and a shadow copy (the main preservation server) that is stored on a Hierarchical Storage Management (HSM) system. A near-site online copy is also preserved on a RAID 5 disc system on a server that is located in another building within the University of Essex. This represents an off-site online copy. A fifth copy is a disc-based offline copy, which is held in either DVD-R or CD-R. All of these copies are formatted according to ISO standards for storage and international standards for housing magnetic and optical media. Mirror versions of on-site systems are also created, and different operating systems are installed across systems to reduce risk (Woollard, 2009).

The policies of the institutions reviewed in this paper make it clear that all files are continuously being monitored and managed by the University of Texas Library, the National Library of Australia and the UK Data Archive. The National Library of Australia remain continuously alert to use of new and appropriate methods for achieving these goals – preferably by means of methods that are fully automated. The UK Data Archive operates media monitoring procedures as part of its AMASS® preservation system. This allows a digital archivist to become aware of any problems that might arise as a result of wear and tear on the media, and to make suitable adjustments

before such problems become too severe. To reduce the incidence of obsolescence, multiple copies of files are stored in different storage media. They are regularly reviewed and are copied into new media whenever this becomes necessary (Woollard, 2009).

The University of Texas Library will soon be using storage media refreshment and file format migration (including the possibility of migration to standard formats during the submission process). File format migration involves moving or converting data from one file format to another file format that is considered to be more stable. The new files that are made in this way are derivative files. Storage media refreshment involves copying data from one long-term storage medium to another of the same type, without any changes being made to the byte stream (binary form of the data).

Information specialists at HKUST see to it that items are migrated to new file formats where necessary. They retain the original byte stream for all items, as well as for any upgraded formats.

The policy document of the National Library of Australia notes that they will use a range of approaches to maintain access to all digital resources, and will then choose the most appropriate approaches to achieve their preservation objectives.

The UK Data Archive provides comprehensive information about their refreshment and migration protocols. Media refreshment and monitoring is overseen during the operation of the archival storage function (as is required by the OAIS Reference Model). Digital Linear Tapes (DLTs) that are used for preservation are re-tensioned every six months, and each full tape in the system is copied every year onto a new tape. These operations are scheduled on an annual basis, and all actions that are performed are carefully logged for the purpose of later checking. Idle tape media are automatically ejected from the DLT drives and are placed in the carousel at set regular intervals. This prevents excessive wear on both the tapes and the drive. The CD-R/DVD-R media are checked according to schedule every two years. If any media reveal either recoverable or non-recoverable errors, they are regenerated from an on-site mirror preservation server. A log is then kept of all the refreshment results, and all the storage media are provided with a date stamp which indicates the time when they were written and the date when they are scheduled for renewal. All CD-Rs are used within three months of purchase. This ensures that only a short amount of time will elapse between the time when they are acquired and the time when they were written (Woollard, 2009). The UK Data Archive follows international best practice in its choice of preservation formats and data migration procedures. When new formats are created from data files – either through migration into new file formats or because of the creation of new file formats for dissemination, the old formats are retained alongside the new (Woollard, 2009).

Automated checks are performed when files are copied onto the preservation system because the migration of files is set up as an automated process. File extensions are always standardized, and a single extension is permitted for each type of file (Woollard, 2009).

All the institutions reviewed in this paper are protected by disaster prevention and recovery plans. Their systems regularly create backups, and, when a repository is closed down, the HKUST, for example, transfers their database to another suitable archive.

All the UK Data Archive servers are protected by power-surge protection systems and carefully considered disaster recovery procedures have been put in place. Woollard (2009) makes reference to the physical safety and security of all their data locations and, among other things, describes their fire prevention and protection systems, their physical intruder prevention and detection systems, and their environmental control systems.

As in all operations of this kind and magnitude, all policies are regularly reviewed and updated. Because of this, most of the custodians of these major digital collections are confident that their practices and procedures are as up-to-date as possible and that they are attuned to whatever new developments are evolving in the world of technology and institutional practice. The University of Texas Library and the National Library of Australia, for example, regularly review their policies. The administrators of the UK Data Archive digital preservation policy annually review their policies and apply version control. They also document control information, provide revision dates and review contact details on an annual basis.

Conclusion

The literature makes it clear that different institutions have adopted different approaches to the practice and planning of digital preservation. Some of these institutions (such as the UK Data Archive) have adopted very comprehensive and detailed preservation policies, while others tend to produce rather more broad and open-ended statements about how they plan to guarantee the accessibility of digital material in the near and remote future. Some of these institutions seem to have no clear understanding of what digital preservation exactly entails. The researcher also observed that while some of the institutions have separate digital preservation policies, others have included them as part of their institution's overall institutional repository policy.

As a result of her investigations, the researcher offers a number of recommendations about issues that need to be addressed in a digital preservation policy. Each of these recommendations is listed under its appropriate heading below:

Content creation

- The use of open file formats should be mandatory since they are openly documented, are supported by a range of software platforms, and are widely used throughout the world. They also have the capacity to address lossless data compression or no compression, are non-proprietary, and do not contain embedded files or embedded programs. (The list issued by the University of Texas Library makes all these advantages clear.)
- Both the original and derivative files should be preserved in all cases.
- Each item submitted should be catalogued with an adequate and sufficient descriptive as well as with all the necessary administrative and structural metadata.
- Quality control should be applied to all processes.
- All sections of a library have something to contribute to the processes of digital preservation, and what they can contribute should be clearly explained.
- All institutions should draw up guidelines that will guide users through the complexities of computer systems.
- All institutions should standardise the use of open standards in their computer systems.

Content integrity

- All preservation policies should comply with the OAIS Reference Model, and all such policies should be constructed in terms of OAIS terminology.
- Preservation policies should be carefully aligned with all the other policies, strategic documents, acts, regulations, standards that apply in an institution.
- All institutions should commission an exhaustive glossary of technical jargon, and this document should be made readily available to all users.
- Submitters need to be issued with a license; the original bitstream should not be altered, and all outstanding problems with regard to rights must be resolved prior to the submission of content.
- All institutions with digital collections should cooperate with other institutions, and keep themselves abreast of all new research and activities in the field by, for example, maintaining relations with the JISC foundation. (http://www.jisc.ac.uk/media/documents/programmes/preservation/jiscpolicy_p1finalreport.pdf)
- Enduring and persistent identifiers should be assigned to each item submitted.
- Each organization should explain how it deals with items that need to be withdrawn from public view.
- The custodians of all digital collections need to address security issues and explain how the content of a repository can become available to authorised users.

- Regular audits of content need to be conducted, and audit trails should be included as part of the functionality offered by the system.
- Each digital archive needs to benchmark itself against the standards and practices of other repositories and their preservation policies in particular.

Content maintenance

- A robust computing and networking infrastructure needs to be kept in place.
- The system that is used should by OAI-PMH compliant.
- The custodians of the archive should adopt a widely used and comprehensive character encoding scheme (Unicode).
- The custodian of the collection should regularly monitor and review the hardware and software that they use, and keep a proven back-up strategy in place.
- Any problems relating to refreshment, migration and emulation should be addressed, and all processes should be automated as far as possible.

The only sure method of ensuring that practices remain current in the context of the development of technological innovation and institutional practices, is to review and update all digital preservation policies on a regular basis.

References

ALCTS Preservation and Reformatting Section (PARS), Working Group on Defining Digital Preservation. (2007). *Definitions of digital preservation* [Online]. ALA Annual Conference, Washington, D.C.June 2007. Available from: http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.cfm. [Accessed: 29 April 2010]

Anderson, C. (2005). Digital preservation: will your files stand the test of time? *Library Hi Tech News* [Online]. 22(6) p.9-10. Available from: http://www.emeraldinsight.com/journals.htm?articleid=1513169&show=abstract. [Accessed: 15 April 2010]

Bevan, S.J. (2007). Developing an institutional repository: Cranfield QUEprints – a case study. *OCLC Systems and Services* [Online]. 23(2) p.170-182. Available from: http://www.emeraldinsight.com/bibliographic_databases.htm?id=1619272&show=abstract. [Accessed: 12 February 2010]

Gibson, H. (2010). *Guidelines for digital preservation* [E-mail]. Message to: I. Smith. 2 June 2010.

Greig, M. & Nixon, W.J. (2007). On the road to Enlighten-ment: establishing an institutional repository service for the University of Glasgow. *OCLC Systems and Services* [Online]. 23(3) p.297-309. Available from: http://sabinet.library.ingentaconnect.com/content/mcb/164/2007/00000023/00000003. [Accessed: 12 February 2010]

Hong Kong University of Science and Technology Library. (2009). *Institutional repository: policies* [Online]. Available from: http://library.ust.hk/info/db/repository-policy.html. [Accessed: 2 June 2010]

Jantz, R. & Giarlo, M.J. (2005). Digital preservation architecture: architecture and technology for trusted digital repositories. *D-Lib Magazine* [Online]. 11(6). Available from: http://www.dlib.org/dlib/june05/jantz/06jantz.html. [Accessed: 12 February 2010]

Jubb, M. (2007). UK funders' policies for the management of information outputs. *The International Journal of Digital Curation* [Online]. 2(1) p.29-48. Available from: http://www.ijdc.net/index.php/ijdc/article/view/36. [Accessed: 29 April 2010]

Lam, K. & Chan, D.L.H. (2007). Building an institutional repository: sharing experiences at the HKUST Library. *OCLC Systems & Services* [Online]. 23(3) p.310-323. Available from:

http://www.emeraldinsight.com/journals.htm?articleid=1622098&show=abstract. [Accessed: 18 April 2010]

McGovern, N. (2009). *DuraSpace digital preservation policies* [Online]. Available from: http://www.fedora-commons.org/confluence/display/FCCWG/Digital+Preservation+Policies. [Accessed: 2 June 2010]

Moore, R. (2008). Towards a theory of digital preservation. *The International Journal of Digital Curation* [Online]. 3(1) p.63-75. Available from: http://www.ijdc.net/index.php/ijdc/article/view/63/82. [Accessed: 27 April 2010]

National Library of Australia. (2008). *Digital preservation policy* [Online]. Available from: http://www.nla.gov.au/policy/digpres.html. [Accessed: 2 June 2010]

Repositories Support Project. (2009). *Digital preservation: are repositories doing enough for preservation?* introduced by Steve Hitchcock [Online]. Tuesday 17 March. Available from: http://www.rsp.ac.uk/podcasts/preservation.php. [Accessed: 27 April 2010]

Strodl, S., Becker, C., Neumayer, R. & Rauber, A. (2007). *How to choose a digital preservation strategy: evaluating a preservation planning procedure* [Online]. Proceedings of the 7th ACM/IEEE-CS International Joint Conference on Digital Libraries, Vancouver, BC, Canada, 2007. Available from: http://portal.acm.org/citation.cfm?id=1255181. [Accessed: 2 June 2010]

University of Texas Libraries. (n.d.). *Digital repository : preservation policy* [Online]. Available from: http://repositories.lib.utexas.edu/policies_preservation. [Accessed: 2 June 2010]

Wise, M., Spiro, L., Henry, G. & Byrd, S. (2007). Expanding roles for the institutional repository. International digital library perspectives. *OCLC Systems and Services* [Online]. 23(2) p.216-223. Available from: http://www.emeraldinsight.com/journals.htm?articleid=1610457&show=html. [Accessed: 27 April 2010]

Woollard, M. (2009). UK Data Archive Preservation Policy [Online]. Available from: http://www.data-archive.ac.uk/news/publications/preservationpolicy.pdf. [Accessed: 2 June 2010]

Addendum A

Table 1: Comparative summary of digital preservation policies

Institution		**University of Texas Libraries**	**The Hong Kong University of Science and Technology Library**	**National Library of Australia**	**UK Data Archive, University of Essex**
Policy content		http://repositories.lib.utexas.edu/policies_preservation	http://library.ust.hk/info/db/repository-policy.html	http://www.nla.gov.au/policy/digpres.html	http://www.data-archive.ac.uk/news/publications/preservationpolicy.pdf
Content creation	Clear and complete technical specifications (such as, for example, those on file formats)	✓			
	Production of reliable master files (the original bitstream)	✓	✓	✓	✓
	Sufficient and adequate descriptive, administrative and structural metadata to ensure future access	✓		✓	✓
	Detailed quality control of processes				✓

Content integrity	Documentation of all policies, strategies and procedures	✓	✓	✓	✓
	Use of persistent and enduring identifiers	✓	✓	✓	✓
	Records of the provenance and stages in the history of all objects	✓	✓		✓
	Verification mechanisms				✓
	Attention to security requirements			✓	✓
	Routine audits	✓		✓	✓
Content maintenance	Robust computing and networking infrastructure	✓	✓	✓	✓
	Storage and synchronization of files at multiple sites	✓	✓	✓	✓

	Continuous monitoring and management of files	✓		✓	✓
	Programs for refreshing, migration and emulation	✓	✓	✓	✓
	Creation and testing of disaster prevention and recovery plans	✓	✓	✓	✓
	Periodic review and updating of policies and procedures	✓		✓	✓