Taming digital records with the Warrior Princess: developing a Xena preservation interface for TRIM

Kye O'Donnell

Kye O'Donnell, is the Senior Project Manager for Recordkeeping Standards at Public Record Office Victoria (PROV). He has recently moved to Melbourne from Perth where he was the Systems Manager for an enterprise content management system at Curtin University of Technology. He has also worked as a records manager for seven years at the City of Perth. He was President of the WA Local Government Records Management Group for five years and is a Business Excellence Evaluator for Standards Australia. He has a Bachelor of Applied Science in Records Management, a Masters in Knowledge Management and is a professional member of the Records Management Association of Australasia and the Australian Society of Archivists.

The preservation of digital records is a challenging task that has been the subject of much debate in the archives and records management community. Following the implementation of an electronic recordkeeping system at the City of Perth, the local government authority of the capital city of the State of Western Australia, a strategy was developed and implemented to ensure that sound processes were established to facilitate the long-term preservation of its digital records. Five major risk areas that affect the preservation of the City's digital records were identified: (1) data format obsolescence; (2) digital conversion errors; (3) loss of original document functionality; (4) storage media obsolescence; and, (5) storage media degradation. To address these risks, the City adopted six digital preservation principles to guide the development of its digital preservation strategy.

The City's digital preservation strategy incorporated the following key elements:

- *using retention and disposal schedules to automatically identify records for digital preservation;*

- *extracting descriptive metadata and electronic records in proprietary data formats from the organisation's electronic recordkeeping system;*

- *normalising digital records using a long-term XML-based preservation data format;*

- *structuring recordkeeping metadata according to a national standard;*

- *encapsulating metadata into normalised records;*

- *human quality assurance processes; and,*

- *storing digitally preserved records online, in the City's electronic recordkeeping system.*

This strategy was implemented through the development of a software interface between the Xena digital preservation tool (produced by the National Archives of Australia) and the TRIM document and records management system (produced by Tower Software). This paper examines the development of the City's digital records preservation strategy and its implementation through a systems integration approach.

Background: the City of Perth

The City of Perth (the City) is the capital city local government of Perth, Western Australia.[1] The geographic area the City currently governs is only 8.1 square kilometres, and includes the Perth central business district and four immediate inner city suburbs.[2] The City is governed by a council of nine members that are democratically elected by the residents and ratepayers located within its boundaries. A chief executive officer reports to the council and is responsible for the operation of the City. The organisation employs 437 permanent, full-time equivalent staff and is structured into four directorates.[3]

In 2000, the City of Perth was using the TRIM document and records management system to manage its paper-based records. Physical files and incoming correspondence were being registered into the system but electronic records were only being stored on shared network drives. In 2001, the City commenced the implementation of an electronic recordkeeping system where records are stored electronically within the City's TRIM system. This system was deployed in four phases, and was fully implemented across the organisation by January 2004. This system supports the business needs of the City, as well as providing recordkeeping functions.

The recordkeeping system (RKS) contains virtual files for classifying and managing electronic records. Scanned hard-copy mail, central email and electronic faxes are collected by the City's Records Services unit and registered and stored in the virtual files in the RKS. Records are also created and received by all staff in the City through their personal email and their PC applications (such as Microsoft Office). All staff capture these records by registering them into the RKS via the system's document management interface.

In accordance with Section 19 of the Western Australian *State Records Act 2000*, the City has a recordkeeping plan that has been approved by the State Records Commission. This document is the primary means of providing evidence of compliance with the State Records Act, and that best practices have been implemented in the organisation.[4] The City's plan was approved on 15 July 2004 for a period of five years.[5] Upon approval the commission requested that a strategy for the ongoing migration of electronic records and data be developed and implemented.

In 2005, the City of Perth developed a strategy for the preservation of its digital records. This document supplements the City's approved recordkeeping plan and was endorsed by the City's executive management team and approved by the State Records Office of Western Australia on 4 July 2005. The objectives of the City of Perth Digital Records Preservation Strategy are to ensure:

- compliance with the State Records Act;
- best practice recordkeeping is conducted in accordance with State Records Commission standards and the Records Management Standard AS 15489;

- sound processes are in place to facilitate the preservation of digital records;

- digital records can be accessed for the duration of their required retention period; and,

- corporate knowledge captured in the City's recordkeeping system can continue to be accessed in the future.[6]

The strategy applies to digital records (including digital images of hard-copy records) of either long-term or archival value stored in the City's recordkeeping system. Preservation activities are applied to the following classes of records, as described in the General Disposal Authority (GDA) for Local Government Records:

- A – records of archival value: all records identified within the GDA with a final disposition of 'A'.

- P – permanent within local government: all records identified within the GDA with a final disposition of 'P'.

- D – destroy by an approved method: all records identified within the GDA with a final disposition of 'D' and a retention period greater than 7 years.[7]

Digital preservation issues

The preservation of digital records is a challenging task that has been the subject of much debate in the archives and records management community. In the text *Keeping Archives* David Roberts identifies two fundamental factors which affect the preservation of digital records: the impermanence of their physical storage media; and, the effects of technological change.[8] Digital preservation literature, such as the digital recordkeeping guidelines issued by the National Archives of Australia, were reviewed to identify preservation issues relevant to the City's records.[9] Five key risks concerning the preservation of the City's digital records were identified, and are briefly discussed below.

Data format obsolescence

The City creates digital records using proprietary data formats, such as Microsoft Word for word-processed documents. The computer software

used to create, edit and view these data formats is the intellectual property of the software companies that produce them. Ongoing access to view these digital records is therefore dependent on the continued availability of this software into the future. Although these data formats are very popular today, the likelihood that the software (and all its versions) will be available and supported on a permanent basis is extremely low.

Digital conversion errors

When converting a digital record from one data format to another, there is always a risk of errors being made during the translation process. If a record is migrated to a new data format each time the organisation upgrades its software, the risk of error increases. The storage of a digital record's metadata (that describes and provides access to the records) in a parallel system can result in the record becoming separated from its metadata during system migrations. This could result in the City being unable to retrieve or understand its digital records.

Loss of original document functionality

The conversion of a digital record to a different data format can lead to a full or partial loss of document functionality. While the new data format may preserve the content and presentation, it may not be possible to easily edit the document to create a new version. Adobe's Portable Document Format (PDF) is a good example of this restriction.

Storage media obsolescence

Computer technology is rapidly changing and improving, resulting in today's technology quickly becoming superseded by new versions and upgrades. If a digital record is stored on out-of-date storage media and the City no longer has the required hardware to read it, the record becomes inaccessible. The retention of superseded computer hardware in a 'computer museum' to read different storage media can only provide a short-term solution, as replacement parts will eventually become unavailable.

Storage media degradation

All digital data is retained on storage media by utilising some physical characteristic of the media to record binary data signals (that is, ones and zeros). Specialised computer hardware is then used to detect (or read) this signal that is stored on the media. Research has shown that over time the physical characteristic of the media will inevitably degrade, resulting in the strength of the signal becoming weaker and weaker until it is finally undetectable. A loss of the signal results in a loss of the digital record.

Digital preservation principles

In order to address all the digital preservation issues outlined above, the City adopted the following set of six principles to guide the development of the City's digital preservation strategy.

Risk-based prioritisation

The preservation of digital records can be a time-consuming exercise and therefore records will be prioritised according to their level of risk. Priority for preservation activities will be given to records of archival value and to formats that are likely to become inaccessible in the short term.

Open source digital data formats

In the publication 'Digital Recordkeeping' the National Archives of Australia states that 'proprietary data formats are not recommended for long-term storage of records', and 'where possible, non-proprietary, fully documented, open source data formats should be used, particularly when implementing migration-based preservation techniques'.[10] Open source data formats are free from intellectual property restrictions and patents, and so anyone can produce software to create, edit or view these digital file formats. Furthermore, the free availability of documentation that describes the data format means that people should have access to the information required to create such software into the future.

Single preservation format

Rather than convert digital records to a new or updated proprietary data format each time the City upgrades its software, a standardised archival data format will be utilised. This will reduce the chance of errors being experienced due to successive migrations over time. It will also be more cost effective in the long term to convert each digital record once to a single preservation format, than to convert all digital records to new data formats with each new software cycle.

Document functionality retention

To ensure it is easy to use and edit digital documents to create new versions, a preservation format that retains the majority of a document's original functionality will be used. The City will also retain the original data format after conversion to a standardised data format. This allows the document's full functionality to be retained, for at least the short term.

Metadata encapsulation

Metadata (or data that describes data) must be captured for digital records to ensure they can be easily retrieved and their content understood. The National Archives of Australia states that metadata 'allows users to control, manage, find, understand and preserve records over time'.[11] The City will adopt a standard set of archival metadata elements (a metadata schema) to simplify the transfer of digital archives to the state archives in the future. The City stores its metadata in its recordkeeping system, with the metadata pointing to digital records in a digital repository that's managed by the system. To prevent the loss of essential metadata regarding digital records, the City will embed (or encapsulate) metadata into its digital records when converting them to an archival data format. This will make the digital records self-describing, assisting with their long-term preservation. The metadata can still be stored in the City's recordkeeping system for retrieval purposes, or to describe records that have been destroyed.

Live storage media

Digital records can be stored in online, offline or near-line storage media. The National Archives of Australia 'strongly recommend that digital records of vital significance to an agency, as well as digital records required for long-term retention within agencies, and digital records of archival value, be stored online.' Online storage systems are routinely maintained and upgraded by the City, reducing the likelihood of storage media degrading or becoming obsolete.

Digital preservation strategy

The City of Perth adopted a nine step strategy, based on the City's digital preservation principles, to ensure the ongoing preservation of its digital records. This strategy is depicted in Figure 1, and each step is described below.
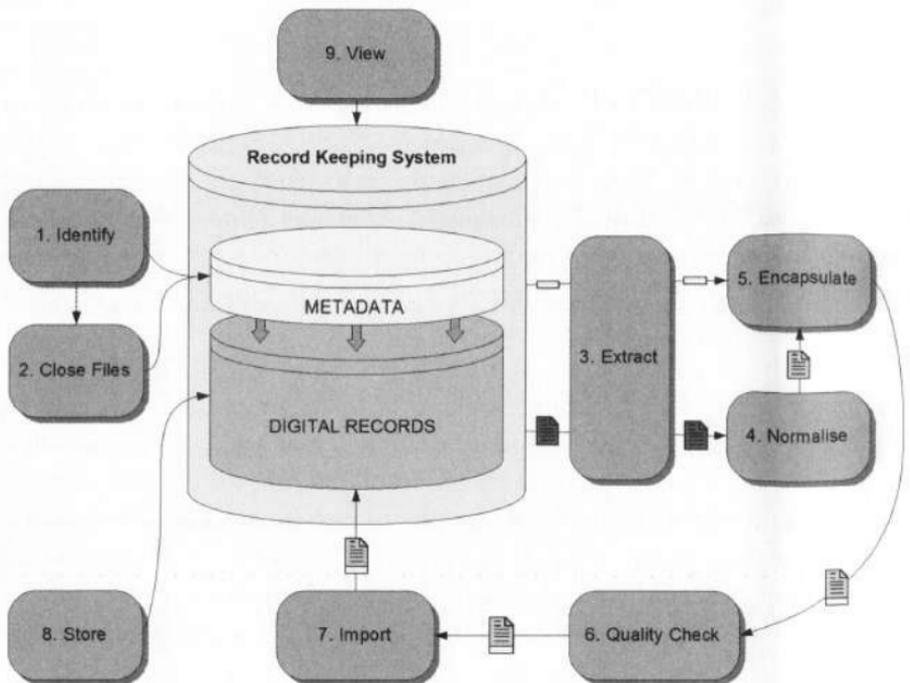

Figure 1. City of Perth Digital Preservation Strategy

Step 1: Identify records for preservation

Digital records of either long-term or archival value are subject to the City's digital records preservation process. As the City has applied the Western Australian General Disposal Authority for Local Government Records to all files in its recordkeeping system, digital records are selected for preservation based on their required retention period. All file volumes with the retention codes listed in Table 1 are scheduled for digital preservation, one year after the date the file volume is closed.

Retention code	Description
A2LA	Archive 2 years after last action: records of archival value to be transferred to the state archives two years after the date of last action.
A5LA	Archive 5 years after last action: records of archival value to be transferred to the state archives five years after the date of last action.
P	Retain permanently in local government: records of permanent value to be retained by the local government.
D8LA	Destroy 8 years after last action: temporary records to be reviewed for destruction eight years after the date of last action.
D9LA	Destroy 9 years after last action: temporary records to be reviewed for destruction nine years after the date of last action.
D10LA	Destroy 10 years after last action: temporary records to be reviewed for destruction ten years after the date of last action.
D20LA	Destroy 20 years after last action: temporary records to be reviewed for destruction twenty years after the date of last action.
D30LA	Destroy 30 years after last action: temporary records to be reviewed for destruction thirty years after the date of last action.
D50LA	Destroy 50 years after last action: temporary records to be reviewed for destruction fifty years after the date of last action.
D71DOB	Destroy 71 years after date of birth: temporary records to be reviewed for destruction seventy one years after the person's date of birth.

Table 1. Retention codes

The above retention schedules in the recordkeeping system are used to automatically flag specific files that are due for digital preservation. As shown in Figure 2, a trigger has been added to the retention schedule's rules so that a file with this schedule applied is scheduled for 'local archive' one year after 'date closed'.
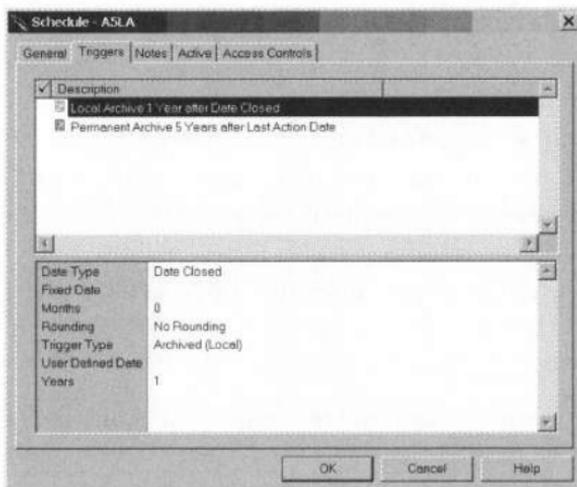
Figure 2. TRIM retention schedule screen

Step 2: Close file volumes

Digital records are only selected for digital preservation after the file volume they are contained in is closed. Each month the City's archives officer closes any file volume that has been open for more than three years and has a retention code listed in Table 1. This ensures that records will be identified for preservation while their data format is still accessible. Relevant files are identified through the use of a 'saved search' as shown in Figure 3. If the file concerns an ongoing issue, a new volume is opened for the capture of new records.
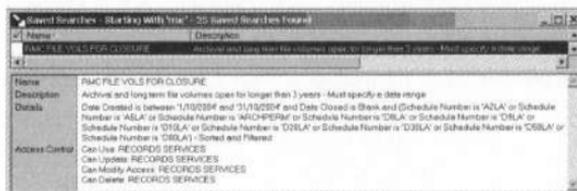
Figure 3. TRIM saved search screen – file volumes for closure

Step 3: Extract original digital records and metadata

Digital records are automatically extracted from the recordkeeping system for preservation when their status is changed to 'archived (local)'. Each month the City's Records Services team reviews a listing of files that are ready for digital preservation and change their status to initiate digital preservation. This is done through the use of another 'saved search' as shown in Figure 4.

Figure 4. TRIM saved search screen – file volumes for archiving

Once a file volume has been flagged for digital preservation, a copy of the final revision of each original digital record contained within the file volume is automatically extracted from the recordkeeping system. Recordkeeping metadata is also extracted for each digital record at this stage.

Step 4: Normalise to standard preservation format

The extracted digital records are then normalised by converting them into a standard preservation data format. The City selected Extensible Markup Language (XML) as its standard preservation data format. XML was chosen because it is an open source format, can be used to store a very wide range of data formats, retains much of the original format's functionality and allows metadata encapsulation. The National Archives of Australia digital preservation software Xena (XML Electronic Normalisation of Archives) is utilised by the City to conduct this normalisation process.

Step 5: Encapsulate metadata in preservation format

The related recordkeeping metadata extracted from the recordkeeping system is structured into a standardised format that complies with the National Archives of Australia *Recordkeeping Metadata Standard for Commonwealth Agencies*.[12] All mandatory and some optional elements of

this metadata schema are met through a combination of static metadata and metadata supplied by the recordkeeping system. The standardised metadata is then encapsulated into the normalised digital record. This effectively makes the digital record self-describing, allowing it to be separated from the recordkeeping system without losing any essential contextual information required to correctly interpret the record.

A mapping between TRIM metadata and the National Archives Recordkeeping Metadata Standard was created so that metadata could be structured according to the standard before it is encapsulated into a digital record. Some metadata is common for all City of Perth records, such as 'corporate name'. This is set as static metadata in the metadata mapping, as shown in Figure 5.

Figure 5. Archivist Client – metadata mapping screen (static metadata)

Most metadata comes from fields in the TRIM database. 'Personal name' for example has been mapped to the 'full name' field in TRIM, as shown in Figure 6. The Archivist Client uses the field names in the TRIM SDK (Software Development Kit) to ensure the mappings are not broken following TRIM software version upgrades.

Some metadata elements require the application of rules to ensure appropriate metadata is captured. The metadata for 'corporate name',

for example, depends on the agent it relates to, as shown in Figure 7. If the agent is a member of the public, then they will not have an organisation listed against them in TRIM and so static text is entered. If the agent is a staff member of an organisation then their organisation's name is used from TRIM.
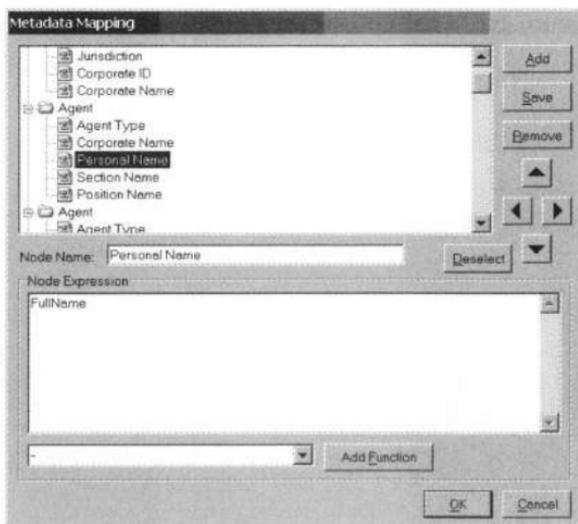
**Figure 6.** Archivist Client – metadata mapping screen (TRIM metadata)
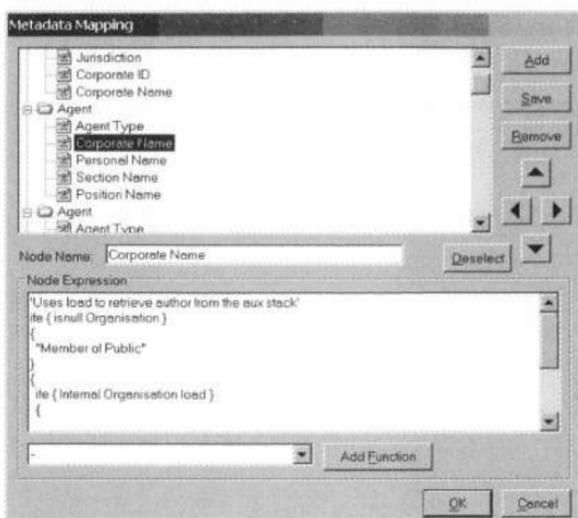
Figure 7. Archivist Client – metadata mapping screen (rule based)

Step 6: Quality check normalised records

Next the normalised records are quality checked by an archives officer. The Archivist Client software (which was developed for the City by Sigma Data Solutions) allows the archives officer to verify the accuracy of the normalisation process by comparing the normalised version against the original format. The main screen of this software lists records that have been normalised, and buttons are provided at the bottom of the window to quality check the records, as shown in Figure 8.

Figure 8. Archivist Client quality assurance screen

Clicking on the 'view Xena' button in the Archivist Client will display a view of the normalised version of the record, as shown in Figure 9.

Figure 9. Xena Viewer – scanned document

Clicking on the 'view original' button in the Archivist Client will display a view of the original version of the record using the TRIM viewer, as shown in Figure 10. This can be used to assess the quality of the normalisation process by visually comparing the two renditions of the same record.

Figure 10. TRIM document viewer – scanned document

When viewing the normalised version of a record, the encapsulated metadata is presented in the bottom portion of the Xena viewer, as shown in Figure 11.
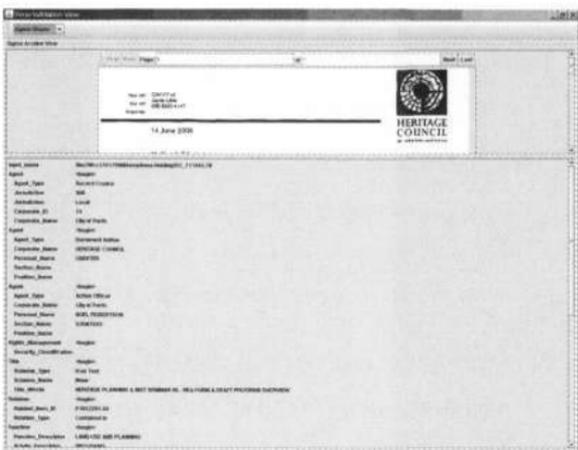
Figure 11. Xena viewer – encapsulated metadata

Clicking on the 'TRIM properties' button in the Archivist Client will display the records metadata stored in TRIM, as shown in Figure 12. This screen can be used to compare against the encapsulated metadata shown in the Xena viewer.
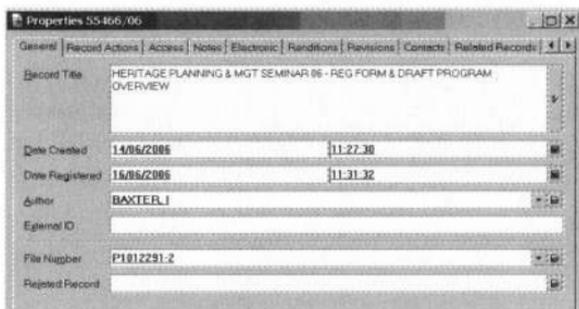
Figure 12. TRIM record properties screen

Step 7: Import preservation format as new rendition

Once a sample of the normalised digital records has been quality checked, the archives officer selects the records and clicks on the 'flag complete' button, as shown in Figure 13. The normalised versions are then imported into the recordkeeping system as a new rendition of the original record. The original digital record is still retained in the recordkeeping system.
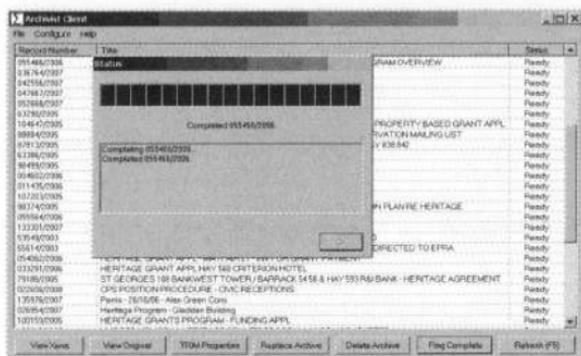
Figure 13. Archivist Client – status screen

The record is flagged as 'finalised' in the recordkeeping system at this stage, as shown in Figure 14, to prevent new revisions being made to the record after it has been preserved. Workflow actions are used in the recordkeeping system to monitor the quality assurance process.

Figure 14. TRIM record details screen

Step 8: Store digital records on live storage media

All digital records that have been digitally preserved are managed by the City's recordkeeping system and stored online, on live digital storage media. Offline storage media is not used to store digital records of long-term or archival value.

Step 9: View preserved records using viewer

Once the normalised digital records are stored in the City's recordkeeping system, they can be displayed for viewing using the Xena archive viewer. To view the normalised version staff access the properties window and click on the 'renditions' tab as shown in Figure 15.

As the TRIM viewer is not capable of viewing Xena files, the TRIM client software must be configured to bypass the TRIM viewer for Xena format files. The Xena viewer must be installed on the desktop computer to be able to view the Xena rendition. Double-clicking on the Xena archive rendition will then open the Xena viewer. Three views of the Xena rendition are available. The XML 'tree view' displays the digital record as raw XML, as shown in Figure 16.

Figure 15. TRIM record properties screen – Xena rendition

Figure 16. Xena viewer – XML tree view

The 'Sigma viewer' option displays a basic, text only view of the digital record with the encapsulated metadata shown at the bottom of the screen, as shown in Figure 17.

Figure 17. Xena viewer – basic view

Clicking on the large button at the top of the basic view displays the digital record using Open Office to provide a correctly formatted view of the record, as shown in Figure 18.


Figure 18. Xena viewer – Open Office view

The view of the digital record in Open Office very accurately represents the appearance of the original record. The same document in Figure 18 is shown in Microsoft Word format for comparative purposes in Figure 19.

Figure 19. Microsoft Word view (for comparison)

To further illustrate the accuracy of the Xena normalisation process, two complex pages from a word-processed document containing tables, bulleted lists, embedded images, line art and coloured text are shown in Figure 20. The images on the left-hand side have been normalised (Xena format) and the images on the right-hand side are in their original proprietary format (Microsoft Word).


Figure 20. Complex documents (Xena left, Microsoft Word right)

Implementation approach

The City's Digital Records Preservation Strategy was implemented in the following six stages over a three-year period.

Stage 1: Modify recordkeeping system retention schedules

The first stage involved modifying the retention schedules listed in Table 1, in the City's recordkeeping system to include a digital preservation trigger. This allows file volumes to be automatically scheduled for digital preservation one year after the closure of the file volume.

Stage 2: Implement file volume closure procedure

Stage two involved the establishment of a new procedure to periodically close relevant file volumes three years after the date the file volumes were opened. As the procedure is performed manually by the City's Records Services staff it was added to the team's end of month checklist to ensure it is routinely completed.

Stage 3: Install and assess Xena

The third stage involved the City downloading and installing a copy of the 'free to use' Xena digital preservation software created by the National Archives of Australia. The software was installed on the TRIM server and selected desktop computers. Xena was used to normalise a sample of digital records created by the City to assess the suitability of the software. All files tested were able to be normalised by Xena and the files that were produced were found to be a very accurate representation of the original record.

Stage 4: Establish export and import process in recordkeeping system

A user-initiated export process was developed for the City by Sigma Data Solutions to extract digital records and their associated metadata from the City's recordkeeping system (TRIM Context). The Archivist Client software was developed to interface with TRIM to extract a copy of digital records and their metadata, pass it to Xena for normalisation and metadata encapsulation, and import the normalised records back into the recordkeeping system as a new rendition to the original record. Xena is designed to accept customised 'plug-ins' to encapsulate metadata into normalised records. A 'plug-in' was created for the City during this stage that structured metadata

according to the requirements of the National Archives of Australia's recordkeeping metadata standard. This process was designed to be as automated as possible, with human intervention only required to select the records for preservation, and quality check the digital records after normalisation.

Stage 5: Test digital preservation process

After the above processes were developed they were rigorously tested in the City's computer testing environment. Testing was conducted using copies of actual digital records kept in the City's recordkeeping system, including a sample of all currently used digital data formats. The normalised digital records were quality checked against the digital records in their original data format to verify that the process preserved the content and presentation of the digital record, as well as the accuracy of the encapsulated metadata. To further test the process a copy of a normalised record was emailed to the National Archives of Australia along with a copy of the City's Xena metadata 'plug-in'. Staff at the National Archives of Australia were able to very easily load the 'plug-in' into their installation of Xena and successfully view the normalised record and its metadata.

Stage 6: Implement digital preservation procedure

Following successful testing, the new digital preservation procedure was implemented on the City's live recordkeeping system. The new procedures were documented and records officers were trained in their operation.

Future directions

The City of Perth's digital records preservation process has been operating successfully for over three years and, thanks to the free availability of Xena, with only minimal cost to the organisation. This has proven that digital preservation software can be successfully integrated with an organisation's recordkeeping system to protect digital records of long-term and archival value while they are still in an organisation's custody. There are however three areas that need further development.

Creating a simple viewer for Xena files

Setting up a computer workstation to be able to view Xena files is unnecessarily complex. The TRIM viewer must be bypassed for Xena files, a Java runtime environment must be operating on the workstation and both Xena (with the metadata 'plug-in') and Open Office software must be installed. Implementing this across a large organisation would be difficult and costly. The development of the TRIM viewer to display Xena files would be the ideal solution. Alternatively a simple executable Xena viewer that could be easily deployed to workstations is required.

Transferring files to the state archives of Western Australia

As well as exporting digital records for preservation, the Archivist Client software has been designed to automatically export normalised digital records when a file's status is set to 'archived (permanent)'. This would allow the City to transfer digital archives to the state archives according to the requirements of the Western Australia General Disposal Authority for Local Government Records. However, the State Records Office of Western Australia is not yet able to accept digital archives into their custody. A pilot project was completed in November 2009 to establish and test a prototype digital archive; this prototype will continue to undergo development throughout 2010–2011.[13] Once a digital repository is established, protocols regarding how digital records should be transferred will need to be developed.

Building preservation into new recordkeeping systems

While the integration between TRIM and Xena is robust, digital preservation processes are essentially operating separate to the recordkeeping system. In the future it would be good to see digital preservation processes being built into document and records management systems so that they are an integral part of the system's retention and disposal activities. Representatives of Tower Software have indicated that they plan to incorporate a digital preservation engine in future versions of TRIM. Curtin University of Technology is also incorporating digital preservation processes into its new enterprise content management system, which uses the open source

software Alfresco.[14] Building digital preservation processes such as that established at the City of Perth into new recordkeeping systems is probably the best method of widely distributing digital preservation technology to organisations.

Endnotes

[1]   City of Perth, *City of Perth information statement 2007–08,* City of Perth, Perth, WA, 2007, p. 3-1.

[2]   ibid., p. 3-2.

[3]   ibid., p. 5-1.

[4]   Government of Western Australia, *State Records Act 2000,* WA State Law Publisher, Perth, WA, 2000.

[5]   Kye O'Donnell, *Record keeping plan,* City of Perth, Perth, WA, 2004, p. 3.

[6]   Kye O'Donnell, *Digital records preservation strategy,* City of Perth, Perth, WA, 2005, p. 3.

[7]   State Records Office of WA, *General disposal authority for local government records,* Library Board of WA, Perth, WA, 1999.

[8]   Judith Ellis (ed.), *Keeping Archives,* second edition, DW Thorpe, Melbourne, 1993.

[9]   National Archives of Australia, *Digital recordkeeping: Guidelines for creating, managing and preserving digital records,* Commonwealth of Australia, Canberra, ACT, 2004.

[10]   ibid., p. 62.

[11]   ibid., p. 31.

[12]   National Archives of Australia, *Recordkeeping metadata standard for Commonwealth agencies,* Commonwealth of Australia, Canberra, ACT, 1999.

[13]   State Records Commission, *Annual Report 2009–2010,* Perth, 2010, p. 13, available at *<http://www.sro.wa.gov.au/src/documents/SRCAnnualReport-0910.pdf>,* accessed 25 October 2010.

[14]   Curtin University, 'CITS Flow Project – Enterprise Content Management for Curtin', webpage available at *<http://flow.curtin.edu.au>,* accessed 18 October 2010.