CANTERBURY DISASTER SALVAGE TEAM

Number 40	**September 2009**
Editor: Lynn Campbell	

NEWSLETTER – Power point presentation from Stephen Clarke, Senior Advisor, Archives NZ. Our key note speaker from the DST workshop "How Safe is Your Hard Copy

Digital Continuity: What is it and How do I develop my own solutions?

Stephen Clarke
Senior Advisor
Digital Sustainability Programme

ARCHIVES NEW ZEALAND
Te Rua Mahara o te Kāwanatanga

The Digital Wild Frontier?

Public sector digital information includes:

- Email
- SMS/Text messages
- Databases
- GIS
- Voice recordings
- Audiovisual recordings
- As well as…

Websites, blogs, wikis, IM, social networking, web 2.0, etc. The www.Wild.Frontier?

Volume

- The volume of digital information being created is increasing exponentially.
- In 2008 the digital content created exceeded storage capacity for the first time.
- By 2011, the volume of digital content will be 10 times the size it was in 2006.
- By 2011, almost half of all information created will not have a permanent home.
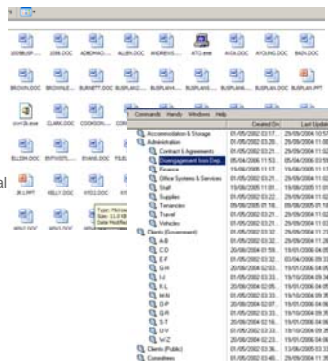
Hardware

- Hardware has a limited life span
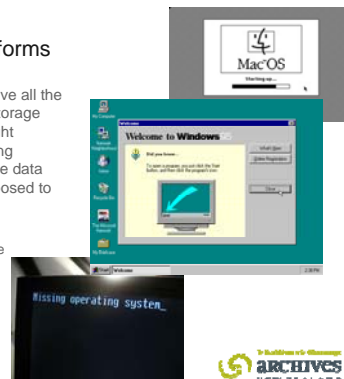
Storage Formats

- Storage has a limited life span

Software Platforms

But assuming you have all the right hardware and storage you then need the right software and operating system to interpret the data and render it as supposed to look.

- Application software
- Operating System
- Display

Media decay

Data corruption (Bit rot)
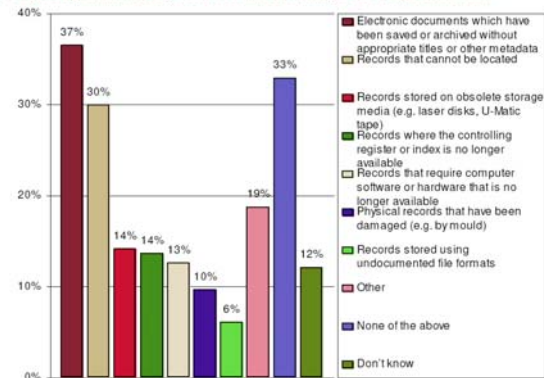
Only one bit of a Byte is corrupted in this image!

Did you know?

- 67% of New Zealand public sector agencies hold records they can no longer access

Electronic Legacy Records Issues (2)

- Proprietary formats and DRM can impact on your ability to access information
- New IT implementations often don't take account of existing systems, information gets orphaned
- Benign neglect is commonplace
- Lack of controlling indexes or context
- Idiosyncratic titling and folder structures
- Lack of organisational awareness and willingness

Graph 4: Q6 Records not in accessible format (n=197)

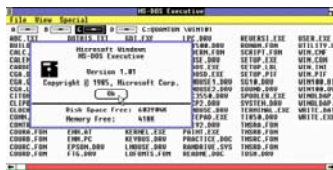*Does your organisation hold any records in a format that means you can no longer access them?*

How do I start?

- Identify what you have
- Make an inventory of formats or software environments you use
- Prioritise 'at risk' information
- Migrate where there are 'quick wins' e.g. from older versions of Microsoft Office products, ppt, Word, Excel, etc.
- Raise awareness and get senior management support
- Draft organisational or departmental policies
- Does the material need to be retained can I dispose?

"Houston, we have a problem"

Make friends with your IT people

Courtesy National Archives of Australia

Electronic Legacy Records Issues

- Electronic records are subject to obsolescence
  - Hardware / media obsolescence.
  - Operating system obsolescence.
  - Software application obsolescence
  - Storage media obsolescence

Steps to Managing e-Legacy Records

- Identify the creators of the records contained in the legacy system
- Identify the physical format
- Determine the software format
- Identify the context of the records' use where possible
- Appraisal to apply, disposal and sentencing, migration strategies and risk analysis
- Convert to open formats

Identifying creators

Implement a institutional knowledge management programme to find out about:
- Organisational administrative history
- Individuals names, roles and positions
- Project working groups
- Previous mergers or amalgamations
- New functions or functions no longer carried out
- What all those %$#@#+# acronyms mean!

Tools that are available to help with identifying file formats include:
- PRONOM
- Droid
- JHOVE
- National Software Reference Library
- Wotsit

Hardware museum

Find out what hardware you have in-house
- 8" Drives, 5 ¼" drives, cartridge players etc.

Find out what software you have in-house
- Earlier versions of windows, Photoshop, in-house developed software, proprietary systems, etc.

Risk Evaluation

- Risk associated with records' formats
- Risk associated with context
- Risks associated with authenticity

Likelihood	Consequences: Extreme	Very High	Medium	Low	Negligible
Almost Certain	Severe	Severe	High	Major	Significant
Likely	Severe	High	Major	Significant	Moderate
Moderate	High	Major	Significant	Moderate	Low
Unlikely	Major	Significant	Moderate	Low	Trivial
Rare	Significant	Moderate	Low	Trivial	Trivial

Table 2: Risk Ranking

- The AS/NZS 4360:1999 Standard on Risk Management
- DRAMBORA
- Trusted Computing

Review

- What is the Business Value?
- Are there Compliance or legal hold considerations?
- Financial implications
  - litigation
  - unnecessary storage costs
  - fraud
  - Loss of contracts or agreements
  - accounts payable/receivable errors and/or omissions

Digital Preservation Tactics

- Normalisation
- Migration (conversion or technology refresh)
- Emulation
- Encapsulation

Open Source Tools

- Fedora – digital archive
- D-Space – digital archive
- DROID – format recognition
- JHOVE – format recognition
- SIARD – database archiving
- XENA – normalisation

- www.sourceforge.net

Open format examples

- ODF - OpenDocument Format
- XML – eXtensible Markup Language
- HTML – Hypertext Markup Language
- PNG – Portable Network Graphics
- FLAC – Free Lossless Audio Codec

- There are emerging mandated and *de facto* standards e.g. PDF(A), OOXML, ODF, JPEG 2000, TIFF, etc.

It's not just a technical issue

- Survey staff on what older e-records they have and encourage them to self migrate
- Use institutional knowledge and find out what systems have been used and where old equipment is
- Engagement is higher when staff feel involved
- Implement policies and procedures so that obsolescence will be managed in future

Popular Myths? (Chris Rusbridge)

1. Digital preservation is very expensive -- not compared to print preservation
2. File formats become obsolete very rapidly -- slower than thought
3. Interventions must occur frequently so costs stay high – less is more
4. Digital preservation repositories should have very long timescale aspirations -- adjust to funding, prepare succession
    - http://www.ariadne.ac.uk/issue46/rusbridge/ (2006)

Popular Myths? (Chris Rusbridge)

1. Digital preservation is very expensive -- not compared to print preservation
2. File formats become obsolete very rapidly -- slower than thought
3. Interventions must occur frequently so costs stay high – less is more
4. Digital preservation repositories should have very long timescale aspirations -- adjust to funding, prepare succession
    - http://www.ariadne.ac.uk/issue46/rusbridge/ (2006)

Digital Continuity at Archives NZ

- **Digital Continuity is ensuring digital information remains available and useable for as long as it is needed**

- Digital Continuity team established 2006
- Digital Continuity Action Plan
- Trialling a Digital Archive for digital archives
- Developing guidance and stuff

Digital Continuity Action Plan – key messages

- **There when you need it**. Public sector digital information will be maintained so that it can be accessed when it is needed.

- **Authentic and reliable**. Public sector digital information is tamper-proof and free of technological digital rights restrictions..

- **Trusted access.** New Zealanders can be confident that they will be able to find, retrieve and use all public sector digital information that can be made publicly available, and that their sensitive information will be protected from unauthorised access.

- **Do nothing, lose everything**. If no action is taken, public sector digital information will be lost.

So what is Archives NZ doing?

- We are developing digital archiving capability both technically and skills
- We have just finished developing the Injestor:
  - Spooler
  - EBCDC
  - Munger
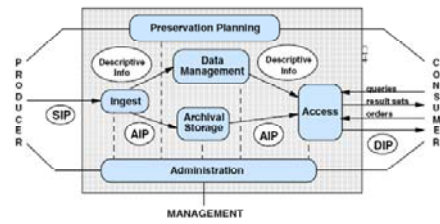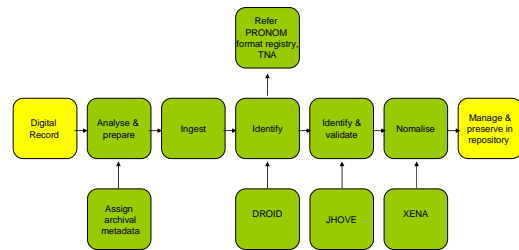 - And other features!

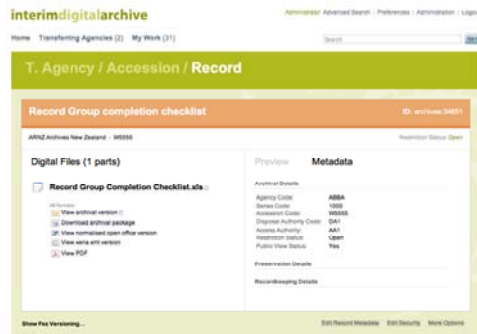## Open Archival Information OAIS

Figure 4-1: OAIS Functional Entities

Interim Digital Archive

How the Interim Digital Archive Works

Ingest and Normalisation

Who to contact

Mick Crouch

Stephen Clarke

New Senior Advisor

Evelyn Wareham

Dougal McLachlan