One man's obsoleteness is another man's innovation.
A risk analysis methodology for digital collections.

Kevin De Vorsey and Peter McKinney
6th May 2009
Assumptions/Assertions

- We do not deny technology (complexity is not necessarily bad)
- We accept all formats
- We do not presume that our risk is the same as others’ (and vice versa)
- We are keeping everything selected for ...(ever?)
NDHA quick facts

No. IEs = 120,731
No. files = 194,573
No. of unique extensions = 428
No. of files with “format unknown” = 5,243
Some rules of engagement

Risk assessment has to be:
- Automated (to a degree)
- Meaningful and obtainable
- Granular
- Cognizant of internal and external factors
- Able to be acted on…(bytestream)
“Obsolescence/obsolete/obsoleteness”

Q1: What does obsolescence/obsolete mean?

Q2: How do we recognise its approach?

Q3: How can we quantify this for analysis?
Defining the Point of Obsolescence

- Format defined
- Early adopters
- Main use phase
- Stopped making software that reads it
- No support for software that reads format
- New version
- Early adopters
- Main use phase
We define format obsoleteness in relation to the Library’s ability to render files within the repository.

If we cannot view, render, or migrate formats then they are “at-risk”.

‘Risk is about the impending loss of the means of providing access’

A PDF is not always a PDF

Interpretation of Standards/Support

Application1

sent to

Application2

interprets

PDF standard

Object 1b

Object 1a
Proposed Solution:

Institutional Libraries that:

1. Ensure the NDHA has a precise understanding of the contents of Rosetta and what degree of it can and cannot be rendered.

And then to:

2. Have a warning system that gives “enough” time to take action to stop files becoming inaccessible.
Library components

• A Local Format Library

• An Application Library
  (that records the Library’s available or tested tools)

• A Risk Library
  (that documents known problems that can affect our ability to render digital objects)
The Libraries Will Document:

- Formats that can be rendered;
- Specific versions of formats that can be rendered;
- The particular characteristics within these versions that are “problematic” (for example compression and colour encoding);
- Applications that can render variations of formats; version and characteristics;
- The sustainability of applications and formats.
Defining the timelines of actions

Our Application library tells us:
An amount of time we know we will have the rendering application for, through:

a. Contract dates with vendor
b. Tech services schedules
c. Controlling the application in the system
d. Vendor support dates
e. Review date if no other date in place.
Application timescales

2004: MacWrite Pro
2006: Photoshop CS2 vendor end support
2008: NOW
2010: Audacity NDHA review
2012: MS End of NLNZ contract
2014: 

PSDs*, TIFFs (except over 2Gb)*
PDF

MP3 -- Fraunhofer*, LAME
WAV -- 24bit, 32 Bit Float
FLAC*

Word 2000
RTF
WPD -- no symbols*
In Summary

- Our risk is based on capability: can we render it?
- This is a relationship between formats and applications
- Characteristics of formats create rendering issues
- We can control our own application destiny (sort of…)

Issue
- We are dependant on tools that are not perfect…therefore automation makes us nervous….