



Te Tari Taiwhenua

RENDERING MATTERS



Report on the results of research into digital object rendering

Euan Cochrane Senior Advisor Digital Continuity Archives New Zealand

Acknowledgements

The contributions to this research of the following individuals and their organisations are greatly appreciated:

- 1. Milton Lee, Digital Preservation Researcher, Victoria University, Wellington, New Zealand.
- 2. Tracey Fenton, Records and Information Manager, SCION, Rotorua, New Zealand.
- 3. Dirk von Suchodoletz, Assistant Professor, Faculty of Engineering, Chair in Communication Systems, The University of Freiburg, Freiburg, Germany.
- 4. Mick Crouch, Advisor/Analyst, Archives New Zealand, The Department of Internal Affairs, Wellington, New Zealand.

Executive Summary

Maintaining the ability of an organisation or user to be able to "open" or "render" a file or set of files is one of the core digital preservation challenges. This report outlines the results of research investigating whether changes are introduced to the information that is presented to users when files are rendered in different hardware and software environments. The report concludes with a set of observations about the impact of the research and provides some recommendations for future research in this area.

Audience

The primary audience for this report is the national and international digital preservation community. This includes any individuals or groups directly involved in planning for, implementing or managing of activities surrounding the preservation of digital objects. The report will likely have a wider audience however. Because of this efforts have been made to explain terms used that are of a technical nature in order to ensure that the barriers to understanding the more technical aspects of this report are minimal.

Secondary audiences for this report include:

- 1. Members of the GLAM sector (Galleries, Libraries, Archives and Museums) who manage or make decisions about the creation, curation, management and preservation of digital objects of long term value.
- 2. The legal community, in particular the electronic-Discovery (eDiscovery) community. The results may help to inform decisions about the role that rendering environments have in regards to the use of digital objects as evidence.
- 3. The wider Information Technology (IT) sector. The research documented in this report highlights the impact that the decisions that application developers have made has had on the ability to preserve information across time. The results may also be used to inform decision making regarding the use (or otherwise) of standard ways of formatting files (standard file formats) for storing information for future access and how software applications are created and maintained over time.

Purpose

The primary purpose of this research is to inform decision making about digital preservation activities. This research involved evaluating how a sample set of digital files opened or "rendered" in different environments (different combinations of software, operating systems and hardware). These files consisted of office administration files ("office files") including word processing document files, presentation files, spreadsheets and databases. The basic methodology involved first opening each file in software that was chosen as the original rendering software and which was running on original hardware from the era. Various aspects of this "control" "rendering" were then documented using a survey tool. These "attributes" included such things as: whether metadata was embedded in the file (normally accessible via a "properties" menu), how images and diagrams were displayed, what word count the software gave, and whether various formatting aspects or fonts were included. Each file was then opened in the same software running on emulated hardware (hardware simulated in software on a host computer) and the same attributes were evaluated to check for any changes. These were documented again using the survey tool. Each file was then systematically opened in a number of modern office software suites and the same attributes were evaluated and the results documented.

The results of this research help to inform digital preservation decision making in a number of ways:

 The results enable digital preservation practitioners to make decisions about the value of ensuring that the original rendering environments are used to open or "render" files in cases where completeness or "authenticity" are important. By comparing different rendering environments to original rendering environments this research helps to show the impact of not using original environments to render digital office files.

- 2. This research will enable digital preservation practitioners to evaluate the use of modern software suites to render old office files and therefore enable practitioners to assess the risk that currently exists to the ability for their agencies to render these office files.
- 3. The research helps to inform decisions about which preservation approaches should be employed to preserve access to complete versions of digital office files, and when each is most appropriate to use. The modern office suites that were tested can be considered migration tools (tools that can be used to move content between files formatted using different format standards). By comparing these to an emulation alternative the results highlight some of the strengths and weaknesses of each approach.

Summary Results

- The choice of rendering environment (software) used to open or "render" an office file invariably has an impact on the information presented through that rendering. When files are rendered in environments that differ from the original then they will often present altered information to the user. In some cases the information presented can differ from the original in ways that may be considered significant.
- The emulated environments, with minimal testing or quality assurance, provided significantly better rendering functionality than the modern office suites. 60-100% of the files rendered using the modern office suites displayed at least one change compared to 22-35% of the files rendered using the emulated hardware and original software.
- 3. In general, the Microsoft Office 2007 suite functioned significantly better as a rendering tool for older office files than either the open source LibreOffice suite or Corel's Word Perfect Office X5 suite.
- 4. Given the effectiveness of modern office applications to open the office files, many files may not need to have content migrated from them at this stage as current applications can render much of the content effectively (and the content's accessibility will not be improved by performing this migration as the same proportion of the content can currently be accessed).
- 5. Users do not often include a lot of problematic attributes in their files but often include at least one. This in turn indicates a level of unpredictability and inconsistency in the occurrence of rendering issues which may make it difficult to test the results of migration actions on files like these.

Contents

Background	9
Outline	9
Scope	9
Methodology	11
Selecting the Control Rendering Environment	11
The Test Environments	12
Original Hardware Test Environments	12
Modern Test Environment	13
Emulated Hardware Test Environments	14
Testing the Rendering of Digital Files	16
Results	18
Summary Statistics	18
Object Rendering	19
Document Rendering	21
Spreadsheet Rendering	24
Database Rendering	27
Presentations and Graphs/Chart Rendering	30
Graph/Chart Rendering	30
Presentation Rendering	31
Findings	33
Modern Rendering Tools and Compatibility/Viability as Migration Tools	33
Modern Software Compatibility with Office Files	33
Modern Software Viability as Migration Tools for Office Files	33
Problematic Attributes	34
Consistency of Rendering Errors	35
Identifying Original Rendering Environments	35
Issues and Opportunites	36
Sample Size	36
Archives New Zealand, The Department of Internal Affairs	5
DMS Library: Archives New Zealand Objective Document ID: A604261	

Comparability	y of Results	36
Software and	Emulated Environment Configuration	36
Future resear	ch	37
Concluding Observa	tions	38
Appendicies		39
Appendix 1.	Attribute Specific Results	39
Documents: /	Attribute Specific Results	39
Spreadsheets	: Attribute Specific Results	43
Databases: A	ttribute Specific Results	47
Appendix 2.	Examples of Problematic Attributes	49
Microsoft Wo	orks 4.0 Document File	49
DBASE Datab	ase File	51
Microsoft Wo	ord for Windows Version 2 Document File	53
WordPerfect	for Windows 5.1 Document File Example 1	55
WordPerfect	Document File Example 2	58
Corel WordPe	erfect version 7 Document File	60
WordStar Do	cument File	64
Appendix 3.	Time required to test the rendering of objects	67
Appendix 4.	Questions in Documentation Questionnaire	68
Appendix 5.	Applications Used in Testing	72

Table of tables

Table 1: Original Test Hardware Configuration	12
Table 2: Modern Test Hardware Configuration	13
Table 3: Emulation Test Hardware Configuration	15
Table 4: Summary results of document rendering tests	22
Table 5: Summary results of spreadsheet rendering tests	25
Table 6: Summary results of database rendering tests	28
Table 7: Attribute specific results for documents	39

Archives New Zealand, The Department of Internal Affairs

Table 8: Attribute specific results for spreadsheets	43
Table 9: Attribute specific results for databases	47
Table 10: Time taken to test x percentage of objects (hours)	67
Table 11: Questions in documentation questionnaire	68
Table 12: Applications used in testing	72

Table of figures

Figure 1: Original Test Hardware	12
Figure 2: Modern Test Hardware	13
Figure 3: Emulated Test Hardware Host PC	14
Figure 4: AQEMU QEMU Graphical Interface	16
Figure 5: Number of Types of Files Tested	18
Figure 6: Migration Tool/Software: Will the object render?	19
Figure 7: Original Software-Emulated Hardware: Will the object render?	20
Figure 8: Percentage of document files tested displaying changes	21
Figure 9: Percentage of attributes displaying changes in at least one tested document file	22
Figure 10: Percentage of spreadsheet files tested displaying changes	24
Figure 11: Percentage of attributes displaying changes in at least one tested spreadsheet file	25
Figure 12: Percentage of database files tested displaying changes	27
Figure 13: Percentage of database attributes tested displaying changes in at least one file	28
Figure 14: "Re-establish links" error message	30
Figure 15: Chart file rendered on emulated hardware	31
Figure 16: Paintbrush slide transition in Freelance Graphics	32
Figure 17: Microsoft Works 4.0 rendering of a document showing word count	49
Figure 18: Microsoft Office 2007 rendering of a Microsoft works 4.0 file with a sentence added and different word count	50
Figure 19: DBASE file rendered in DBASE IV showing table structure	51
Figure 20: DBASE file rendered in Microsoft Access 2007 showing table structure	52
Archives New Zealand, The Department of Internal Affairs	7
Divide Library a Archives New Zealand Objective Document ID: A004201	

Figure 21: Word for Windows version 2.0 file rendered in Word 6.0c for Windows 3.x	53
Figure 22: Word for Windows version 2 file rendered in LibreOffice Writer version 3.3.0	54
Figure 23: WordPerfect for Windows 5.1 file rendered in Word Perfect for Windows 5.2	55
Figure 24: WordPerfect for Windows 5.1 file rendered in WordPerfect for Windows 6.1	56
Figure 25: WordPerfect for Windows 5.1 file rendered in Microsoft Office 2007	57
Figure 26: WordPerfect file rendered in WordPerfect 5.2 for Windows	58
Figure 27: WordPerfect file rendered in Microsoft Office 2007	59
Figure 28: Corel WordPerfect version 7 file rendered in Corel WordPerfect version 7	60
Figure 29: Corel WordPerfect version 7 file rendered in LibreOffice Writer 3.3.0	61
Figure 30: Corel WordPerfect version 7 file rendered in Microsoft Office 2007	62
Figure 31: Corel WordPerfect version 7 file rendered in Corel WordPerfect X5	63
Figure 32: WordStar file rendered in WordStar version 7 for MS-DOS	64
Figure 33: WordStar file rendered in Microsoft Word 5.5 for MS-DOS	65
Figure 34: WordStar file rendered in LibreOffice Writer 3.3.0	66

BACKGROUND

Archives New Zealand anticipates receiving for preservation a large quantity of office administration files (word processing documents, spreadsheets, presentations and databases) over coming decades. Upon receiving these records Archives New Zealand needs to make decisions about how best to preserve and provide access to them now and in the future. The research documented in this report is intended to inform decision making by providing examples of the efficacy and impact of different approaches that are currently available.

OUTLINE

This report contains the results of research testing the rendering of office files in different digital environments, conducted over two months at Archives New Zealand in 2011. The results give an indication of the effectiveness of currently available tools for use in preserving access to content in digital objects and access to complete versions of the digital objects.

A limited analysis of the results of this research is included in this report in the "Findings" section. The Findings section also includes a discussion of the not-insignificant issues with the research and a discussion of the options available for mitigating these in future research.

The report concludes with a set of observations about the impact of the research and provides some recommendations for future research in this area.

SCOPE

The research documented in this report was conducted over a period of approximately 9 months in 2011 at Archives New Zealand. During the research 110 files had their rendering tested (i.e. the information that was presented and the way the information was presented when opened in different environments). The testing component of the research was completed over approximately 2 months. Each file was tested in an average of 4.7 environments including the original software running on original hardware, the original software running on emulated hardware and three different modern office suites, Microsoft Office 2007, LibreOffice version 3.3.0 and Corel WordPerfect Office X5. Discussion of the scoping decisions relating to the selection of the rendering applications is included in the methodology section.

The number of files tested was limited due to the time and resources that were available. Each rendering test took on average 9 minutes to complete for a total of nearly 80 hours for all 523 renderings (one incomplete). According to the metadata embedded in the files in the sample, they were last modified during the time period 1988-2002. The National Archives (TNA)'s DROID tool¹ was able to identify 12 formats amongst the files and numerous versions of each² format. The file formats as reported by Droid included:

• 1 xlc file Microsoft Chart File

¹ More details on DROID and PRONOM are available here: <u>http://sourceforge.net/apps/mediawiki/droid/index.php?title=Main_Page</u> (accessed 12/12/11)

² The DROID results for each file are included in the dataset that will be published alongside this report.

Archives New Zealand, The Department of Internal Affairs

- 5 unknown: 3 .wdp WordPerfect files and two .mdb database files
- 37 .doc files with multiple matches (most likely Microsoft word files)
- 22 .xls workbook files of at least two different formats
- 3 .wps OLE2 files
- 1 Quattro pro spreadsheet file
- 9 .dbf dBase files
- 2 .pre Freelance Graphics presentation files
- 7 Framework database files
- 1 .pub publisher file
- 6 .db Paradox database files
- 16 .doc/.wp6 WordPerfect files.

METHODOLOGY

The procedure followed when conducting this research was as follows:

- 1. Files were selected for testing
- 2. A control rendering environment was established
- 3. The file was opened in the control environment and its attributes³ documented using a survey tool
- 4. The file was opened in a modern office suite and changes to its attributes documented using a survey tool
- 5. Step 4 was repeated for the 3 modern office suites
- 6. The file was opened in the emulated hardware environment and changes to its attributes documented using a survey tool
- 7. The testing process progress was logged in a spreadsheet.

SELECTING THE CONTROL RENDERING ENVIRONMENT

In order to have something to which to compare modern office suites and emulated hardware environment renderings a control rendering environment had first to be established. If the original rendering software was known this software was selected. This was possible for the files that were provided from personal collections and for files that only ever had one creating application (e.g. Framework II files).

Where the original rendering software was not known the following selection method was followed:

- 1. The age of the selected file was used to determine which possible operating system it may have been created on.
- 2. The database of software applications was then searched to find applications running on the operating systems that correlated to the age of the file.
- 3. This subset of applications was searched to find those that created files with the file extension of the selected file and would also open them.
- 4. The DROID output identifying the format of the file (and through inference the default rendering application) was compared against these applications to ascertain whether any of them matched.
- 5. The selected file was opened in each identified application to attempt to discover whether it obviously rendered "properly" in one of them.

³ Various aspects "rendering" of the files were documented. These "attributes" included such things as: whether metadata was embedded in the file (normally accessible via a "properties" menu), how images and diagrams were displayed, what word count the software gave, and whether various formatting aspects or fonts were included.

Archives New Zealand, The Department of Internal Affairs

6. If no obvious differences were identified when opening the file in multiple applications then the most common application, and/or the application that previously created files of the determined type by default, (rather than as a selectable option) was chosen as the control application.

THE TEST ENVIRONMENTS



ORIGINAL HARDWARE TEST ENVIRONMENTS

Figure 1: Original Test Hardware

Four test environments were set up with operating systems and software from their era installed on them. The four computers were also attached to a Keyboard-Video Monitor-Mouse (KVM) switch (5.). This enabled all four machines to be used with the same keyboard, monitor and mouse.

Details of the hardware configuration of the test systems are outlined in the table below:

Table 1: Original Test Hardware Configuration

#	Operating System	Model	Processor	Ram	Video	Audio	Network
1	MS-DOS 6.22 & Windows 3.11 for Workgroups	Compaq ProLinea MT 4/66	486/DX Upgraded to Pentium 83 Overdrive	16 MB	Cirrus Logic GD-5Uxx VGA	ESS Audio MPU-401	3Com Etherlink III
2	Windows 95	Compaq Deskpro 5100	Pentium (r) 100MHz	24 MB	Compaq QVision PCI v1.2.0.3A	Sound Blaster 16 (CT4170)	AMD PCNet

Archives New Zealand, The Department of Internal Affairs

3	Windows 98 SE	Compaq Deskpro EP/SB	Pentium 350/100 Mhz	256 MB	NVidia Vanta/Vanta LT	Creative Soundblaster Live!	3Com 3c905B-TX
4	Windows 2000	Compaq Evo W4000	Pentium 4 1.7 Ghz AT/AT compatible	786 MB Ram	NVidia Geforce 2 MX/MX 400	SoundMax	Intel Pro/100

MODERN TEST ENVIRONMENT



Figure 2: Modern Test Hardware

The Modern Test Environment consisted of a somewhat aged Pentium 4 HT 3 Ghz machine with Windows 7 Professional Installed on it.

The specifications of the machine are outlined below:

Table 2: Modern Test Hardware Configuration

Operating System	Model	Processor	Ram	Video	Audio	Network
Windows 7 Professional	HP Compaq DC7600 Small Form Factor	Pentium 4 3 Ghz with Hyper Threading	1.5 GB	Intel® Graphics Media Accelerator 950	Integrated High Definition audio with Realtek 2 channel ALC260 codec	Intel 82945G Express

Three modern office suites were installed on Windows 7. The three office suites were:

1. LibreOffice version 3.3.0 OOO330m19 (Build:6)

Archives New Zealand, The Department of Internal Affairs

DMS Library: Archives New Zealand Objective Document ID: A604261

- 2. Corel WordPerfect Office x5 version 15.0.0.357
- 3. Microsoft Office Professional 2007

Only stand-alone modern office suites that were available for installation on Microsoft Windows 7, were under active development and which had on-going support options were selected for the experiments.

In order to limit the scope of the experiments the following decisions were made regarding applications to be used for testing the rendering for files in the experiments:

- 1. Only one office suite based on the OpenOffice.org program code was selected (LibreOffice based on Go-oo, an OpenOffice fork).
- 2. No cloud based office suites were selected.
- 3. OS X products were not included in the experiments.
- 4. Corel Office was not included due to the inclusion of Corel WordPerfect Office (a different product from the same vendor).

Lotus SmartSuite was not included in the test set as it is no longer being actively developed and was not available for the Microsoft Windows 7 operating system.

Microsoft Office 2010 was not selected for this test because it had only recently been released when the research was being planned at the end of 2010.



EMULATED HARDWARE TEST ENVIRONMENTS

Figure 3: Emulated Test Hardware Host PC

The emulated hardware test environments were hosted on a small modern laptop with a Pentium Core2 Duo Processor and 2 GB of RAM. The host system was running Ubuntu Linux version 11.04 and had QEMU version 0.14.0 installed on it with the AQEMU Graphical User Interface (GUI) application installed to help with configuration and execution of the emulated environments.

Table 3: Emulation Test Hardware Configuration

Operating System	Processor	Ram	Video	Audio	Network
MS-DOS 6.22 & Windows 3.11 for Workgroups	IBM PC 32 Bit	24 MB	StdVGA (VESA 2.0)	Sound Blaster 16	Default (E1000)
Windows 95	IBM PC 32 Bit	256 MB	Cirrus CLGD 5446	Sound Blaster 16	None
Windows 98 SE	IBM PC 32 Bit	256 MB	Default (Cirrus CLGD 5446)	Sound Blaster 16	None
Windows 2000	IBM PC 32 Bit	256 MB	Cirrus CLGD 5446	Sound Blaster 16	Default (E1000)

A reasonably novel approach was used in order to ensure that all software dependencies that were included in the original software environments, running on the original hardware, were included on the software environments on the emulated hardware. Images were made of the hard drives in the original hardware environments and these images were then attached to emulated hardware. The hard drive images then had to have some changes made so that they would be compatible with the emulated hardware (such as installation of new drivers). After this the images were able to be booted from on the emulated hardware with identical sets of installed applications as on the original hardware. This approach ensured a great deal of similarity between the original software running on the original hardware environment and the original software running on the emulated hardware environment⁴.

QEMU/AQEMU EMULATION SOFTWARE AND INTERFACE

The emulation software selected for use in this research was the open source emulator QEMU. QEMU is a true emulator in that it does not rely on virtualization (running code directly on the underlying hardware of the host system). Instead QEMU emulates the hardware needed in software on the host computer and runs the desired operating system and application code on top of that emulated hardware. "When used as a machine emulator, QEMU can run OSes [Operating Systems] and programs made for one machine (e.g. an ARM board) on a different machine (e.g. your own PC). By using dynamic translation, it achieves very good performance."⁵

QEMU is controlled by the command line by default. In order to simplify the configuration and use of QEMU a Graphical User Interface application called AQEMU was used. AQEMU enables easy selection of the emulated hardware and attaching of various disk images containing the software to be run on the emulated hardware, and the

⁴ This process was described in detail in the paper "Replicating Installed Application and Information Environments onto Emulated or Virtualized Hardware by Suchodoletz, D. and Cochrane, E. presented at the iPRES 2011 conference: <u>http://ipres2011.sg/</u>

⁵ <u>http://qemu.org/</u>

Archives New Zealand, The Department of Internal Affairs

files to be tested within that software. Once configured, emulated environments can be booted in QEMU using AQEMU with a click of the "play" button:

	🕹 🎯 🔄 🔹 🌣 AQEMU			💷 🔿 🕸 🖂 16:41 😣 digital 😃
Framework2 Autoload Demo	Info General Device Manager Network	Ports Advanced Other		
	Machine Name: Win 311 + MSDOS 6.22		Boot Priority:	HDD 🛟 🚱
Win 311 + MSDOS 6.22	Emulator Type: QEMU	÷	<u>C</u> PU Type:	Default 🗘
Nin98Corel	Emulator Version: Default	* *	Number of CP <u>U</u> :	1
	Computer Type: IBM PC 32Bit	* *	Video Card:	StdVGA (VESA 2.0)
Windows 2000	Machine Type: Default	*	<u>K</u> eyboard Layout	en-us (default)
Windows 311 and DOS 6.22				
	Memory Size			
Windows 95		1 I I I I I		24 MB 🔻
Nindows 98	1 MB	🕝 Free memory: 2571	MB	2768 MB
	Remove limitation on maximum amo	unt of memory		
	Audio Cards Sound Blaster 16	ENSO	NIQ ES1370	Yamaha YM3812
	□ AC97 (Intel 82801AA)	Gravis	Ultrasound GF1	PC Speaker
	Fullscreen Mode	📈 Snar	shot Mode	Vse Local Time
		— ,	_ ~	
				✓ <u>Apply</u> Discard Changes
💼 🔅 AQEMU				9

Figure 4: AQEMU QEMU Graphical Interface

QEMU has the ability to run the emulated environments in "snapshot mode". This mode enables full interaction with the emulated environment but does not save any changes that are made. This was found to be very beneficial during this research as it enabled the provision of identical environments every time they were needed for testing and it also enabled the technician to be trained without worry of altering the emulated environments.

TESTING THE RENDERING OF DIGITAL FILES

Having selected the control environment the technician then opened the file in that environment and proceeded to answer a series of questions (a subset of 136 potential questions) about the content presented through the rendering of the file using the control environment.

A full list of the questions used is included in Appendix 4. The questions were presented to the technician using the open source LimeSurvey Survey Software tool⁶. This enabled question routing to be used and enforced consistency in the questions answered by requiring them to be answered before enabling the survey to progress. The survey will be made available alongside this report in LimeSurvey XML format and queXML Survey format for reuse and interrogation purposes.

After completing the questionnaire with the control environment the technician then opened the file in one of the migration or emulation environments and proceeded through the questionnaire again from the beginning. Automatic

⁶ <u>http://www.limesurvey.org/</u>

question routing was used to ensure the questions presented by the questionnaire were relevant for the type of file being tested and the type of environment it was being tested within.

Progress through the test set was documented in a spreadsheet, the data from which was included with the final results data.

RESULTS

The results of the research are presented below in tabular and graphical formats without comment. Attribute-specific results are included in Appendix 1. The full dataset included detailed results for all files will be released alongside this report.

SUMMARY STATISTICS



Figure 5: Number of Types of Files Tested

Figure 5 shows the total number of each type of file that was tested. The sample set included more document files than any other type.

OBJECT RENDERING



Figure 6: Migration Tool/Software: Will the object render?

Figure 6 shows the total number of times files were opened in migration tools/Modern office application suites and how often they rendered at all and how often they rendered with and without errors. This data includes multiple results for each file as each file had their rendering tested in multiple modern office suites.

The difference between the "Partially" and "With Errors" categories was considered ambiguous and therefore they can be combined for evaluation purposes but were displayed in these results for transparency.



Figure 7: Original Software-Emulated Hardware: Will the object render?

Figure 7 shows the number of times files were able to be rendered when they were opened in original software running on emulated hardware.

DOCUMENT RENDERING



Figure 8: Percentage of document files tested displaying changes

Figure 8 shows the percentage of all the document files that were tested with each environment that displayed at least one change from the original rendering when rendered using the alternative environment.



Figure 9: Percentage of attributes displaying changes in at least one tested document file

Observations of changes to a predefined set of attributes were carried out for each alternative test environment that was included in the research. Figure 9 shows a generalised result across all the files that were tested. For each alternative rendering environment, it shows the percentage of attributes for which at least one file displayed a change in that attribute. For example, 86% of attributes displayed a change in at least one file that was tested with LibreOffice.

The particular file that displayed a change in any one attribute in any one environment may have been different for each attribute (i.e. " file A" may have had a change in metadata when rendered in LibreOffice Writer and "file B" a change in word count when rendered in the same environment etc).

Table 4: Summary results of document rendering tests

	Microsoft Word 2007	Corel WordPerfect X5	LibreOffice Writer	Original Software on Emulated Hardware
Number of files tested	55	51	56	54
Number of files with changes	33 (60%)	51 (100%)	56 (100%)	12 (22%)
Number of attributes	17 (59%)	22 (76%)	25 (86%)	7 (24%)

Archives New Zealand, The Department of Internal Affairs

displaying changes in at least one file				
Number of files with changes (excluding metadata ⁷)	32 (58%)	49 (96%)	56 (100%)	10 (19%)
Number of files the application could open (with or without errors)	55/55 (100%)	51/55 (93%)	56/56 (100%)	54/54 (100%)

⁷ Metadata issues potentially introduced by the use of incorrect file copying procedures when setting up the equipment initially seemed to be biasing the results. This result was included in order to attempt to account for that potential issue.

Archives New Zealand, The Department of Internal Affairs

SPREADSHEET RENDERING



Figure 10: Percentage of spreadsheet files tested displaying changes

Figure 10 shows the percentage of all the spreadsheet files that were tested with each environment that displayed at least one change from the original rendering when rendered using the alternative environment.



Figure 11: Percentage of attributes displaying changes in at least one tested spreadsheet file

Observations of changes to a predefined set of attributes were carried out for each alternative test environment that was included in the research. Figure 11 shows a generalised result across all the files that were tested. For each alternative rendering environment, it shows the percentage of attributes for which at least one file displayed a change in that attribute. For example, 62% of attributes displayed a change in at least one file that was tested with LibreOffice.

The particular file that displayed a change in any one attribute in any one environment may have been different for each attribute (i.e. " file A" may have had a change in metadata when rendered in LibreOffice Calc and "file B" a change in word count when rendered in the same environment etc).

Table 5: Summary results of spreadsheet rendering tests

	Microsoft Excel 2007	Corel Quattro Pro X5	LibreOffice Calc V3.3.0	Original Software on Emulated Hardware
Number of files tested	24	21	21	28
Number of files with changes	14 (67%)	20 (95%)	21 (100%)	9 (32%)

Archives New Zealand, The Department of Internal Affairs

Number of tested attributes displaying changes in at least one file	21 (66%)	19 (58%)	20 (62%)	3 (9%)
Number of files the application could open (with or without errors)	24/25 (96%)	21/28 (75%)	21/28 (75%)	28/28 (100%)

DATABASE RENDERING



Figure 12: Percentage of database files tested displaying changes

Figure 12 shows the percentage of all the database files that were tested with each environment that displayed at least one change from the original rendering when rendered using the alternative environment.



Figure 13: Percentage of database attributes tested displaying changes in at least one file

Observations of changes to a predefined set of attributes were carried out for each alternative test environment that was included in the research. Figure 13 shows a generalised result across all the files that were tested. For each alternative rendering environment, it shows the percentage of attributes for which at least one file displayed a change in that attribute. For example, 75% of attributes displayed a change in at least one file that was tested with LibreOffice.

The particular file that displayed a change in any one attribute in any one environment may have been different for each attribute (i.e. " file A" may have had a change in metadata when rendered in LibreOffice Base and "file B" a change in word count when rendered in the same environment etc).

Table 6: Summary results of database rendering tests

	Microsoft Access 2007	Corel Quattro Pro X5	LibreOffice Base v3.3.0	Original software on Emulated Hardware
Number of files tested	9	6	11	17
Number of files with changes	8 (89%)	9 (69%)	10 (91%)	6 (35%)

Archives New Zealand, The Department of Internal Affairs

Number of tested attributes displaying changes in at least one file	6 (35%)	4 (67%)	10 (75%)	4 (22%)
Number of files the application could open (with or without errors)	9/14 (64%)	6/12 (50%)	11/16 (69%)	17/17 (100%)

PRESENTATIONS AND GRAPHS/CHART RENDERING

Two presentation files and one graph file were included in the test set. These low numbers mean that no generalisations can be made about them. However, some interesting observations were able to be made about them and these are included below.

GRAPH/CHART RENDERING

The graph file was a Microsoft Excel Chart file from 1993 with an ".xlc" file extension. These files are not supported by Microsoft Office 2007, LibreOffice version 3.3.0 or Corel WordPerfect Office X5. Microsoft Office 2003 will open the files and is still supported by Microsoft. However recent updates to Microsoft Office 2003 (Service Pack 3) remove support for this format and other older formats for security reasons⁸. Support for handling older formats can be re-enabled in Office 2003 if users are comfortable with the security implications but this requires either downloading and running an executable file or editing Windows registry settings.

The graph rendered with very few errors when using the original software running on the emulated hardware. There was a perceptible difference in colour in the image on screen. This was likely due to poor calibration of the computer hosting the emulator providing the emulated hardware.

LINKED FILES

In both the original hardware rendering and the emulated hardware rendering a message was presented when the file was opened that asked the user "This document contains links. Re-establish Links?":

Microsoft Excel - EXGRAPH.XLC	- +
Eile Edit View Insert Format Tools Window Help	\$
□ ■	
Microsoft Excel Microsoft Excel This document contains links. Re-establish links? Yes No Help	
Figure 14: "Re-establish links" error message	

⁸ <u>http://support.microsoft.com/kb/938810</u> (accessed on 16/11/11)

The file containing the data that was linked to in this chart was in the same folder as the chart file and so it was easy to re-establish the link by selecting the appropriate file. After this the file opened with no other problems:



Figure 15: Chart file rendered on emulated hardware

This example highlights the need for a tool to enable digital preservation practitioners to be able to identify which files have dependencies like this when selecting files for ingest. If practitioners could identify such files then they could take steps to ensure that the dependencies can be preserved alongside the primary digital files that require them.

PRESENTATION RENDERING

The two presentation files tested in these experiments were Freelance Graphics Files from 1994 with ".pre" file extensions. These files are not supported by Microsoft Office 2007 or LibreOffice version 3.3.0. Corel WordPerfect Office X5 was able to open the presentations however it introduced significant changes when doing so. Although a version of Freelance Graphics for Windows XP is still supported by IBM⁹, it was not included in these experiments for reasons already mentioned. It may have been able to render the files but its ability to do so was not tested.

The slides rendered with very few errors when running on the original software running on the emulated hardware. There was a slightly perceptible difference in colour in the slides as presented on screen and some slight image

⁹ http://www-01.ibm.com/support/docview.wss?rs=899&uid=swg21253536 (Accessed on 16/11/11)

Archives New Zealand, The Department of Internal Affairs

distortion issues. This may have been due to poor calibration and general testing of the computer hosting the emulator providing the emulated hardware and the configuration of the emulation software.

The Freelance Graphics application included a number of innovations including unique slide transitions such as the paintbrush transition:

V2QEMU ("Windows 311")	<u> </u>
Example Organisat Chart	
CEO	
manager	
pleb pleb pleb	
super pleb super pleb	

Figure 16: Paintbrush slide transition in Freelance Graphics

These transitions were presented equally well in the original as well as when run on the emulated hardware environment. Audio included in one presentation also rendered well via the emulated hardware.

MODERN RENDERING TOOLS AND COMPATIBILITY/VIABILITY AS MIGRATION TOOLS

MODERN SOFTWARE COMPATIBILITY WITH OFFICE FILES

None of the tools provided full backwards compatible renderings of all older file types. The Microsoft Office 2007 suite of applications displayed substantially better results when rendering older files than either LibreOffice version 3.3.0 or Corel WordPerfect Office X5. This finding is interesting when compared to the claims made by the marketer of Corel's product. Corel's marketing for WordPerfect Office X5 claims that it has: "market-leading compatibility"¹⁰. This claim may be true of modern file formats but the results outlined in this report bring into question the compatibility of the product with older files. The standout compatibility of Microsoft's product may be due to a large number of the test files being structured according to Microsoft formats (up to 60%) and likely having been created in Microsoft applications as discussed further below.

In general the original software running on the emulated hardware was able to render the objects included in the test set much more consistently than the modern office suites. Only 22-35% of the files rendered using the emulated hardware and original software had any changes compared to the 60-100% across the modern office suites.

The emulated original software environments were also able to open every single file (with or without immediate errors) that was tested with them. Unfortunately every single file was not opened in every single modern application so it is not possible to accurately compare the ability for each environment to open each file from the data that was produced by these experiments. Rectifying this issue is a goal for future research. The modern office suites failed at being able to open the test files at all (with or without errors) a number of times each for both spreadsheets and databases and LibreOffice and WordPerfect Office also failed a few times for the document files they were tested with.

MODERN SOFTWARE VIABILITY AS MIGRATION TOOLS FOR OFFICE FILES

A common approach to the preservation of office files is using tools like the XENA tool from the National Archives of Australia either to normalise or migrate the content from the old files to new files formatted in an open format.

There are a number of steps involved in the use of modern office suites as migration tools. Practitioners first need to open the digital files. In an automated migration process this is done by the migration tool (e.g. LibreOffice). Secondly they need to save the old files as new files formatted according to a more recent, or sustainable, formatting standard. These files can then be opened in modern software which is usually intended to be the same tool that was used for migration. The viability and effectiveness of this process depends on a number of factors. One of these factors is the ability of the modern software to render the newly created file content the same way it did when it was rendered from the file formatted using the pre-migration formatting standard. Assuming the modern software doesn't introduce any additional change when rendering the new file format, then the results of this research can be assumed to provide an evaluation of the effectiveness of each of the modern office suites as migration tools by showing how they render the content post-migration.

¹⁰ See <u>http://www.corel.com/content/vpk/wpox5/Corel_WordPerfect_Office_X5_Reviewers_Guide.pdf</u> page 2 (accessed on 15/11/11)

Archives New Zealand, The Department of Internal Affairs

Given this caveat, some conclusions can be drawn about the ability for the modern office suites to be used as migration tools and their relative effectiveness at preserving the information presented to the users without changes when compared to using emulation:

- 1. In general, the Microsoft Office 2007 suite functions significantly better as a migration tool than either the open source LibreOffice suite or Corel's Word Perfect Office X5 suite.
- 2. The emulated environments, with minimal testing or quality assurance, provide significantly better preservation functionality than the modern office suites when the modern office suites are used as either rendering tools or migration and subsequent rendering tools.
- 3. Given the effectiveness of modern office applications to open the office files, many files do not need to have content migrated from them at this stage as current applications can render much of the content effectively (and the content's accessibility will not be improved by performing this migration as the same proportion of the content can currently be accessed).

PROBLEMATIC ATTRIBUTES

Some attributes were more problematic (displayed rendering errors more frequently) than others. The most common rendering errors were as follows:

For Documents:

- Changes to embedded metadata (e.g. the "document author" metadata displayed in the "properties" menu changed)
- Information added (e.g. a sentence was visible in a modern software rendering that was not visible in the original rendering)
- Number of words reported (e.g. 234 became 254)
- Content positioning/pagination (e.g. paragraphs were placed at different positions on the page and the page count changed)
- Fonts (e.g. Times New Roman became Courier New, etc)

For Spreadsheets:

- Changes to embedded metadata (e.g. the "file creator" metadata displayed in the "properties" menu changed)
- Rounding (e.g. the number of decimal places displayed changed or the way rounding was calculated was changed to use a different algorithm)
- Fonts (e.g. Times New Roman became Courier New, etc)

For Databases:

- Changes to embedded metadata (e.g. the "file creator" metadata displayed in the "properties" menu changed)
- Information added (e.g. fields contained information where they did not previously)
- Internal structure changes (e.g. links between fields were removed and/or field restrictions were changed)

Archives New Zealand, The Department of Internal Affairs

Documents had a larger numbers of files showing rendering issues with more attributes than either spreadsheets or databases. This may have been due to the larger sample set. Examples illustrating some of the problematic attributes, including screenshots, are included in Appendix 2.

CONSISTENCY OF RENDERING ERRORS

The results outlined above demonstrate that both:

a) The likelihood that any single file has a particular attribute that does not render properly in a particular rendering environment is low,

and

b) The likelihood that the same file will have at least one attribute that doesn't render properly in a particular environment is quite high (~60% and above).

In other words, the results indicate that users do not often include a lot of attributes in their files that caused rendering issues when rendered in modern environments but often include at least one. This in turn indicates a level of unpredictability and inconsistency in the occurrence of rendering issues.

A significant challenge for digital preservation practitioners is evaluating the effectiveness of digital preservation approaches. When faced with a large and ever increasing volume of digital files to be preserved, practitioners are forced to consider approaches that can be automated. The results in this report indicate that the occurrence of problematic attributes is inconsistent and they therefore may be difficult to automatically identify. Without identifying such attributes pre-migration it will not be possible to test whether the attributes exist post-migration and so the effectiveness of the migration will not be able to be evaluated. Without *automatically* identifying such attributes pre-migration then it is unlikely that any effective evaluation will be able to be made cost-effectively. The cost to manually identify these attributes for every object would likely be prohibitively large for most organisations given reasonably sized collections.

IDENTIFYING ORIGINAL RENDERING ENVIRONMENTS

Information identifying the original rendering environments of digital objects has at least two uses in digital preservation:

- 1. It can be used to test migration actions to confirm that all the content is still there by (for example) opening the file in the original rendering environment and in the migrated environment to compare across the two.
- 2. If emulation is to be used as a preservation approach, information about the rendering environment is required in order to identify which environment needs to be preserved for use in rendering digital objects in the future.

Identifying the original rendering environment (as required for this research) was found to be difficult in this project. The information that DROID produced was often not very useful. There were multiple format matches for many files and a number that did not have any matches. Even when DROID produced an exact format match it could only be taken as a strong indicator that the application that was usually associated with that format was the one that created the file. There was at least one file in the set that was almost definitely not created by the application that was normally associated with the format that was matched to it by DROID¹¹. This finding highlights the need for more work on tools to identifying creating applications and for pro-active documentation of rendering environments by content creators. In the past, website creators regularly stated on their sites which browser(s) their site rendered best in. This sort of information would be very valuable if it was available for office files and would then negate the need for tools to automatically identify such information so long as the provided information could be trusted.

ISSUES AND OPPORTUNITES

There were a number of issues with the research documented in this report that should be seriously considered when evaluating the results. In order to provide transparency and to mitigate some of these issues, the raw data produced by this research will be made available alongside this report. Archives New Zealand will also seek to make available the files that were used for the testing as well where possible¹². Both of these actions will help to enable others to replicate the research and confirm the results for themselves.

SAMPLE SIZE

A significant issue with this research is the size of the sample set. The small numbers of spreadsheet and database files tested (in particular) make the conclusions drawn from the results of this research unreliable at best. It would be worthwhile to extend these experiments to cover a larger number of files.

A further reason why a larger sample set would be beneficial is that many attributes were not apparent in any of the files in the test set and so no data was able to be collected about them. Extending the research to a larger sample set would hopefully give more examples of more varied files enabling more comprehensive testing to be completed.

COMPARABILITY OF RESULTS

The number of files tested with each application in each category often differs (e.g. the number of documents tested with Word 2007 differs from those tested with LibreOffice Writer). This shows that some files were not tested in every application but only in some (or their tests were not documented). This makes the comparisons between results from rendering files in different applications less reliable than they could be. For future research it should be ensured that every file that is tested in one environment is tested in every environment. The data produced by this project could also be reanalysed to exclude those files for which results are not available for every appropriate rendering environment.

This issue also means that the results may not show how often files failed to open at all when tested in some applications. Instances of these failures may not have been recorded. This means that the statistics on how often files could be opened with or without errors may be inaccurate.

SOFTWARE AND EMULATED ENVIRONMENT CONFIGURATION

At least two caveats need to be placed on the results of this research due to inadequate procedures used when configuring the different test environments.

¹¹ There were two copies of a report in the set, one (obviously the original due to rendering details) seemed to be created in WordPerfect 7 and another that was identified by DROID as a Microsoft Word file had a .doc extension and appeared to have been created by WordPerfect 7 also.

¹² Some files contain information of a confidential or personal nature and will not be able to be released.
While analysing the data, issues became apparent with the way the test sample set of files was copied for use in the different test environment. Some of the results relating to the changes to embedded metadata when displayed in the emulated and migrated environments may be due to the metadata having inadvertently been updated in the process of copying the files for use in the other environments.

More thorough comparisons between the original and emulated environments should also have been conducted to test the configuration of the emulator providing the emulated environment. The configuration used in the testing was the result of little to no testing. A default emulated environment setup was used with compatible drivers installed for the operating system so it could run on the changed hardware. This may have meant that some attributes of the original rendering were changed when files were rendered on the emulated hardware environment such as the colours displayed on screen.

FUTURE RESEARCH

Due to resource and time constraints only a limited amount of analysis was conducted on the data produced by this research. A number of additional analyses could be conducted based on the data produced by this research. These include:

- a) Analysis of rendering efficacy by "File format".
- b) Analysis of rendering efficacy by vendor/suite.
- c) Analysis of rendering efficacy by age of file.
- d) Analysis of the occurrence of problematic attributes in relation to the occurrence of other problematic attributes.

Analysis of the results has shown that it is not always clear that the emulated/migrated questions were answered properly e.g. that a file displayed attribute x in the original environment rendering but not in the modern software rendering. It would be worth comparing the data on the migration-tool environment and emulated hardware environment renderings to the control renderings to ascertain how accurately the results were captured.

It would be very useful if other researchers were to take the framework established by this research and apply it to a larger set of files and to different types of digital objects. For example, a similar set of research into the rendering of archived websites would be very beneficial. Replicating this research to test its results would also be very beneficial to ensure that the conclusions that may be drawn from the results are justified and to remove some of the issues that were identified with the approach used in this research.

CONCLUDING OBSERVATIONS

The results of this research are interesting and valuable for a number of reasons:

They show quantitatively that the choice of rendering environment (software) used to open or "render" an office file invariably has an impact on the information presented through that rendering. When files are rendered in environments that differ from the original they will often present altered information to the user. In some cases the information presented can differ from the original in ways that may be considered significant. This result is useful as it gives a set of ground-truth data to refer to when discussing the impact of rendering on issues of authenticity, completeness and the evidential value of digital office files.

The results give an indication of the efficacy of modern office suites as rendering tools for older office files. Risk analysis of digital objects in current digital repositories could be informed by this research. Digital preservation risk analysts could use this research to evaluate whether having access to these modern office suites means that files that can be "opened "by them are not at risk.

The results highlight the difficulty and expense in testing migration approaches by showing how long it took to test only ~100 files comprehensively (at least 13.5 hours). Scaling this to 0.5% of 1,000,000 files would give 675 hours or nearly 17 weeks at 40 hours per week. This level of testing may be considered excessive depending on the context, but similarly comprehensive testing of only 100 files per 1,000,000 of each format (.01%) would take at least 13.5 hours per format, per tool. More information on how long testing would take for a variety of different sample sizes and percentages of objects (e.g. 1% of 100,000 objects would take 150 hours) is available in Appendix 3.

The results also show the promise of running original software on emulated hardware to authenticate the rendering of files to ensure that all the content has been preserved. Although emulated environment renderings were not shown to be 100% accurate in this research, they were shown to have a far greater degree of accuracy in their renderings than current office suites (which are the tools currently used for migrating office files). Additionally, some of the changes introduced in the emulated environments may have been due to poor environment configuration.

The results give an indication of how prevalent certain attributes are in office files. With a greater sample size this research this could help to show whether or not it is true that "most users only use the same 10% of functionality in office applications"¹³ (the data from this small sample indicates that in fact they only use about 10% of the functionality/attributes each, but often it is a different 10%).

The research will generally be very beneficial in informing the decisions made by Archives New Zealand regarding the preservation of digital office files. By releasing the results data and survey used in the research alongside this report it is hoped that the wider digital preservation community can gain value from this research also and ideally either replicate or extend the research to a wider sample set and/or different object domain.

¹³ See: <u>http://blogs.technet.com/b/john_westworth/archive/2008/03/06/but-my-users-only-use-10.aspx</u> (last accessed on 7/12/11)

APPENDIX 1. ATTRIBUTE SPECIFIC RESULTS

Detailed results were compiled for each object type that was reported on. These results include information on how many files showed changes in each environment for each attribute that was tested.

DOCUMENTS: ATTRIBUTE SPECIFIC RESULTS

Table 7: Attribute specific results for documents

	Number of document files that had changes to attributes when rendered in environments other than the control								
Attribute	Microsoft Word 2007	Corel WordPerfect X5	LibreOffice Writer v3.3.0	Original software (Emulated Hardware)					
Embedded Metadata ¹⁴	33 (60%) 21 (41%) 38 (68%)		38 (68%)	1 (2%)					
Fonts	20 (36%)	17 (33%)	25 (45%)	0 (0%)					
Formatted text	0 (0%)	1 (2%)	2 (4%)	0 (0%)					
Highlighted text	0 (0%)	0/2 that included highlighted text (0%)	0/2 that included highlighted text (0%)	0/3 (0%) that included highlighted text					
Coloured text	0/1 that included coloured text (0%)	0/1 that included coloured text (0%)	0/1 that included coloured text (0%)	0/1 (0%) that included coloured text					
Page dimensions	6 (11%)	4 (8%)	20 (36%)	0 (0%)					
Pagination	17 (31%)	11 (22%)	33 (59%)	4 (7%)					

¹⁴ May be due to the way the sample was copied and made available in the emulated environment.

Archives New Zealand, The Department of Internal Affairs

	Number of document environments other t	files that had changes to han the control	o attributes when rende	ered in	
Attribute	Microsoft Word 2007	oft Word Corel WordPerfect LibreOffice Write X5 v3.3.0		Original software (Emulated Hardware)	
Position of content on page/screen	22 (40%)	20 (39%)	39 (70%)	4 (7%)	
Line spacing	4 (7%)	1 (2%)	19 (34%)	0 (0%)	
New-line placement	5 (9%)	1 (2%)	21 (38%)	0 (0%)	
Page and/or section break placement	17 (31%)	17 (33%)	36 (64%)	4 (7%)	
Orientation of objects or text	8 (15%)	2 (4%)	14 (27%)	0 (0%)	
Justification of text	3 (5%)	2 (4%)	17 (30%)	0 (0%)	
Information added	38 (69%)	9 (18%)	21 (38%)	4 (7%)	
Custom shapes	3/23 that included custom shapes (13%)	5/22 that included custom shapes (23%)	3/13 that included custom shapes (23%)	0 (0%)	
Embedded Images	1/5 that included embedded images (20%)	¾ that included embedded images (75%)	½ that included embedded images (50%)	¼ that included embedded images (25%)	
Hidden Content	0/17 that included hidden content (0%)	4/8 that included hidden content (50%)	½ that included hidden content (50%)	2/8 that included hidden content (25%)	
Watermarks	0/9 that included watermarks (0%)	3/9 that included watermarks (33%)	1/7 that included watermarks (14%)	0/9 that included watermarks (0%)	

	Number of document files that had changes to attributes when rendered in environments other than the control							
Attribute	Microsoft Word 2007	Corel WordPerfect X5	Original software (Emulated Hardware)					
Custom Character sets	3/8 that included custom character sets (38%)	1/5 that included custom character sets (20%)	2/5 that included custom character sets (40%)	0/6 that included custom character sets (0%)				
Custom languages or language interfaces	0/1 that included custom languages or language interfaces (0%)	1/1 that included custom languages or language interfaces sets (100%)	1/1 that included custom language or language interfaces sets (100%)	0/1 that included custom languages or language interfaces (0%)				
Number of words reported by software	25 (45%)	44 (86%)	51 (91%)	0 (0%)				
Footnotes or endnotes	0/27 that included footnotes or endnotes (0%)	5/28 that included footnotes or endnotes (18%)	3/22 that included footnotes or endnotes (14%)	0 (0%)				
Current date object	0/3 that included current date objects (0%)	0/4 that included current date objects (0%)	2/4 that included current date objects (50%)	0/4 that included current date objects (0%)				
Lists or bullet points	0/10 that included lists or bullet points (0%)	0/17 that included lists or bullet points (0%)	0/17 that included lists or bullet points (0%)	0/16 that included lists or bullet points (0%)				
List or bullet point symbols	2/10 that included lists or bullet points (20%)	1/17 that included lists or bullet points (6%)	2/17 that included lists or bullet points (12%)	0/16 that included lists or bullet points (0%)				

	Number of document environments other t	The files that had changes to attributes when rendered in than the controlCorel WordPerfect X5LibreOffice Writer v3.3.0Original 						
Attribute	Microsoft Word 2007	ber of document files that had changes to attributes when render comments other than the controlbisoft WordCorel WordPerfect X5LibreOffice Writer v3.3.0that included 						
Borders	0/14 that included borders (0%)	0/15 that included borders (0%)	1/9 that included borders (11%)	0/15 that included borders (0%)				
Citations	1/3 that included 0/2 that included citations (33%) citations (0%)		1/3 that included citations (33%)	0/3 that included citations (0%)				
Mail-merge settings	0/1 that included mail-merge settings (0%)	1/1 that included mail-merge settings (100%)	1/1 that included mail-merge settings (100%)	0/1 that included mail- merge settings (0%)				
Links to other files	0/2 that included links to other files (0%)	0/0 that included links to other files (0%)	0/2 that included links to other files (0%)	0/2 that included links to other files (0%)				

SPREADSHEETS: ATTRIBUTE SPECIFIC RESULTS

Table 8: Attribute specific results for spreadsheets

	Number of spreadsheet files that had changes to attributes when rendered in environments other than the control							
Attribute	Microsoft Excel 2007	icrosoft Excel Corel Quattro Pro X5 LibreOffice Calc 07 v3.3.0						
Embedded Metadata	19 (79%)	19 (90%)	21 (100%)	9 (32%) ¹⁵				
Links to Other Files	0/1 that included links to other files (0%)	0/1 that included links to other files (0%)	1/1 that included links to other files (100%)	0/1 that included links to other files (0%)				
Macros/Scripts	0/0	0/0	0/0	0/1 that included macros or scripts (0%)				
Fonts	5 (21%)	8 (38%)	5 (24%)	1 (4%)				
Formatted text	0 (0%)	1 (5%)	0 (0%)	0 (0%)				
Highlighted text	0/1 that included highlighted (0%)	0/1 that included highlighted (0%)	0/1 that included highlighted (0%)	0/1 (0%)				
Page dimensions	5 (21%)	3 (14%)	2 (10%)	0 (0%)				
Pagination	6 (25%)	4 (19%)	3 (14%)	0 (0%)				
Position of content on page/screen	5 (21%)	3 (14%)	2 (10%)	0 (0%)				

_

¹⁵ May be due to the way the sample was copied and made available in the emulated environment.

Archives New Zealand, The Department of Internal Affairs

	Number of spreadshe environments other t	Corel Quattro Pro X5 LibreOffice Calc v3.3.0 Original software (Emulated Hardware) 1 (5%) 1 (5%) 0 (0%) 0 (0%) 0 (0%) 0 (0%) 4 (19%) 2 (10%) 0 (0%) 0 (0%) 0 (0%) 0 (0%) 5 (24%) 1 (5%) 0 (0%) 3 (14%) 6 (29%) 0 (0%)					
Attribute	Microsoft Excel 2007	Corel Quattro Pro X5	LibreOffice Calc v3.3.0	Original software (Emulated Hardware)			
Line spacing	5 (21%)	1 (5%)	1 (5%)	0 (0%)			
New-line placement	5 (21%)	0 (0%)	0 (0%)	0 (0%)			
Page and/or section break placement	6 (25%)	4 (19%)	2 (10%)	0 (0%)			
Orientation of objects or text	5 (21%)	0 (0%)	0 (0%)	0 (0%)			
Justification of text	6 (25%) ¹⁶	5 (24%)	1 (5%)	0 (0%)			
Information added	5 (21%)	3 (14%)	6 (29%)	0 (0%)			
Custom shapes	0 (0%)	1/1 that included custom shapes (100%)	0 0(%)	0/1 that included custom shapes (0%)			
Embedded Images	0/1 that included embedded images (0%)	2/2 that included imbedded images (100%)	1/2 that included imbedded images (50%)	0/2 (0%)			
Hidden Content	1/1 that included hidden content (100%)	0/4 that included hidden content (0%)	0/2 that included hidden content (0%)	0/2 that included hidden content (0%)			
Formulae	0/16 that included formulae (0%)	0/15 that included formulae (0%)	1/16 that included formulae (6%)	0/23 that included			

¹⁶ Included adding a named range to one file.

Archives New Zealand, The Department of Internal Affairs

	Number of spreadsheet files that had changes to attributes when rendered in environments other than the control							
Attribute	Microsoft Excel 2007	Corel Quattro Pro X5	LibreOffice Calc v3.3.0	Original software (Emulated Hardware)				
				formulae				
Formulae Notation	1/16 that included formulae (6%)	15/15 that included formulae (100%)	2/16 that included formulae (13%)	0/23 that included formulae				
Rounding calculation method	5 (21%)	1 (5%)	4 (19%)	0 (0%)				
Number of decimal places displayed	5 (21%)	1 (5%)	5 (24%)	0 (0%)				
Internal links	0/3 that included internal links (0%)	1/2 that included internal links (50%)	0/2 that included internal links (0%)	0/7 that included internal links (0%)				
Current-Date Object	0/1 that included a current date object (0%)	0/1 (0%)	0/1 that included a current date object (0%)	0/1 that included a current date object (0%)				
Coloured Cells	0/6 that included coloured cells (0%)	0/5 (0%)	0/6 that included coloured cells (0%)	0/6 that included coloured cells (0%)				
Cells with Borders	0/8 that included cells with borders (0%)	0/10 (0%)	2/7 that included cells with borders (29%)	0/10 that included cells with borders (0%)				
Conditional Formatting	0/1 that included conditional formatting (0%)	0/1 that included conditional formatting (0%)	0/1 that included conditional formatting (0%)	0/2 that included conditional formatting				

	Number of spreadshe environments other t	et files that had changes han the control	s to attributes when ren	dered in
Attribute	Microsoft Excel 2007	Corel Quattro Pro X5	LibreOffice Calc v3.3.0	Original software (Emulated Hardware)
				(0%)
Column order	5 (21%)	0 (0%)	0 (0%)	1 (4%)
Row order	5 (21%)	0 (0%)	0 (0%)	0 (0%)
Functions	4 (21%)	1 (5%) 2 (10%)		0 (0%)
Named ranges	0/4 that included named ranges (0%)	0/2 that included named ranges	2/3 that included named ranges (66%)	0/3 that included named ranges (0%)
Cell types	5 (21%)	4 (19%)	0 (0%)	0 (0%)
Links to other spreadsheets	0/1 that included links to other spreadsheets (0%)	0/1 that included links to other spreadsheets (0%)	1/1 that included links to other spreadsheets (100%)	0/2 that included links to other spreadsheets (0%)
Multiple worksheets	0/8 that included multiple worksheets (0%)	0/12 that included multiple worksheets (0%)	0/9 that included multiple worksheets (0%)	0/9 that included multiple worksheets (0%)

DATABASES: ATTRIBUTE SPECIFIC RESULTS

Table 9: Attribute specific results for databases

	Number of database files that had changes to attributes when rendered in environments other than the control							
Attribute	Microsoft Access 2007	LibreOffice Base v3.3.0	Original software (Emulated Hardware)					
Embedded Metadata	7 (78%)	3 (50%)	0 (0%)	0 (0%)				
Links to Other Files	1/1 that included links to other files (100%)	0/0 that included links to other files (0%)	0/0 that included links to other files (0%)	2/3 (67%) ¹⁷				
Macros/Scripts	1/1 that included Macros or Scripts (100%)	0/0 that included links to other files (0%)	0/0 that included links to other files (0%)	0/2 (0%) ¹⁸				
Fonts	0 (0%)	1 (17%)	1 (9%)	2 (12%)				
Coloured text	2/2 that included coloured text (100%)	1/1 that included coloured text (100%)	1/1 that included coloured text (100%)	0/1 that included coloured text (0%)				
Page dimensions	0 (0%)	1 (17%)	1 (9%)	0 (0%)				
Pagination	0 (0%)	1 (17%)	1 (9%)	0 (0%)				
Position of content on page/screen	0 (0%)	1 (17%)	1 (9%)	0 (0%)				

¹⁷ Linked files missing

¹⁸ Could be executed but did not work due to missing linked file.

Archives New Zealand, The Department of Internal Affairs

Line spacing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
New-line placement	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Page and/or section break placement	0 (0%)	1 (17%)	1 (9%)	0 (0%)
Orientation of objects or text	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Justification of text	0 (0%)	0 (0%)	9 (82%)	0 (0%)
Information added	1 (11%)	5 (83%) 2 (18%)		1 (6%)
Custom front- end/interface	0/1 that included custom front-ends or interfaces (0%)	0/0 that included 0/0 that included custom front-ends or custom front- interfaces (0%) or interfaces (0/2 that included custom front- ends or interfaces (0%)
Saved queries	0/1 that included saved queries (0%)	0/0 that included saved queries (0%)	0/1 that included saved queries (0%)	0/2 that included saved queries (0%)
Internal structure (e.g. primary keys, table links etc).	4 (44%)	5 (83%)	2 (18%)	3 (18%)
-				

APPENDIX 2. EXAMPLES OF PROBLEMATIC ATTRIBUTES

MICROSOFT WORKS 4.0 DOCUMENT FILE

This research did not include any attempt to evaluate the value or importance of any of the attributes that were tested. To help to enable these judgements to be made some examples of changes that were identified have been included in the section below.

Microsoft Works- Fred and Maggie - Tim speech.wps _ 8 × 🐑 File Edit View Insert Format Iools Window Help _ 8 × Times New Roman 🝷 12 🝷 🗗 🚍 🛃 👌 👗 🖻 健 🗷 s me nens n It's just another sparrow's dream . Above a London street But I'll remember all my life A bird called Marguerite." Microsoft Works X There are 334 words in your document. **i**) Works has counted the words in your document, including footnotes and header/footer paragraphs. OK 1?> + Counting words: Pg 2/2 Start Nicrosoft Works - [Fr... 📢 🚺 1:08 PM

Figure 17: Microsoft Works 4.0 rendering of a document showing word count

	🚽 🔊 - 1) ⇒			Fred an	nd Maggie - 1	Tim speech.wps	- Microsoft W	Vord			
0	Home	Insert	Page Layout	References	Mailings	Review	View					0
	Home	Insert	Page Layout	References 1 · 2 · 1 · 3 · 1 id, "mine's id, "the me we I'll alway ter taking varrows fell e up to the did, "those ain't the there's the ust another a London st ll remember called May id locked r 90 { Q l € :kkkkkkkkkkkkk E ^L · L · L Z o] Š :	Fred an Mailings 4 · 1 · 5 · 3 Marguer Emory of Arguer and bread an L asleep 2 sound of bells of 2 bells of 2 sparrow treet 3 all my rguerite. my hands 2 S Å ckkkkkk 	Ad Maggie - Review 1 · 6 · 1 · 7 rite" our day" nd wine of bells know" f Notre 1 of Bow" of Bow" of Bow" of Bow" of a drear life " about h A ê - J b yyyyyyyyyy yyy yyy yyyyyyyyy gg A XTimes 1 A ê = E	<pre>im speech.wps View '' Dame" a is neck. '' Dame" a is neck. '' '' New '' '' '' '' '' '' '' '' '' '' '' '' ''</pre>	The stick	c swished L □ '1 ≪1 S ^J ≪1 S ^J ≪1 8• Z• □ Y	d twice a: T Ž T J L T L T L T L T L L L L L L L L L L L L L	s 	
												* 0
Page: 10	of 10 We	ords: 477	English (New 7)	alandì						<u> </u>		
			1 🛷					1			EN 🔺 🎼 🛅	4:57 p.m.

Figure 18: Microsoft Office 2007 rendering of a Microsoft works 4.0 file with a sentence added and different word count

Figures 17 and 18 show how the word count of the Microsoft Works 4.0 file went from 334 to 477 words when rendered in Microsoft Word 2007 and a sentence fragment is now made visible that was not displayed anywhere in the original. The sentence fragment reads: "mbed and locked my hands about his neck. The stick swished twice as th". This content may have been previously deleted in the original but continued to be stored somewhere in the file.

Figure 18 also shows that extra symbols have been added to the bottom of the document. These symbols are likely formatting information that has been interpreted as text by Microsoft Word 2007.

The formatting and font in the original rendering of the document have also changed. The centre-alignment of the poem in the Microsoft Works 4.0 rendering has changed to left-alignment in the Microsoft Word 2007 rendering and the font changed from Times New Roman to Courier New in the Microsoft Word 2007 rendering.

DBASE DATABASE FILE

Layout	t Organize	Append Go	To Ex	it		1:2:	1:46 pm
						. Butes nemaining'	3929
Num	Field Name	Field Type	Width	Dec	Index	bytes remaining.	3565
1	SCESTMON	Numeric	2	0	N		
Z	SKEY	Character	12		N		
3	CKEY	Character	7		N		
4	SCCROPNO	Numeric	1	0	N		
5	SCESTNO	Numeric	1	0	N		
6	SCESTSPH	Numeric	5	0	N		
7	SCESTMET	Character	1	2000000	N		
8	SCROTATN	Numeric	1	0	N		
9	SCSITEIN	Numeric	4	1	N		
10	SCGMODEL	Character	3		N		
11	SCHTAGE	Character	3		N		
12	SCUSER1	Character	1		N		
13	SCUSER2	Character	1		N		
14	SCUSER3	Character	1		N		
15	SCUSER4	Character	1		N		
16	SCLASTME	Character	4		N		
		<u>e — </u>					
Databas	se D:Nsc525	5-10\SCROP	Field	1/22			
	Enter tl	he field name	. Insert.	/Delete	e field:(Ctrl-N/Ctrl-U	
Field r	names begin w:	ith a letter a	and may o	contaiı	n letter:	s, digits and under:	scores

Figure 19: DBASE file rendered in DBASE IV showing table structure

	. .	(? • • •					Table Too	SCROP : Database ((Access 2007)	- Microsoft Access		
0	Home	Create	Externa	al Data	Database T	ools	Design					0
View Views	Primary Key	Builder Tes	st Validation Rules Tools	Great G	isert Rows elete Rows ookup Column	Prope	erty Indexe et	25				
All Table	es			sc	ROP							×
SCROP			*		Field N	lame		Data Type			Description	
AB SC	ROP			SCE	STMON			Number				
ub				SKE	EY			Text				
			1	CKF	FY			Text				
			7	SCC	ROPNO			Number				
			7	SCE	ESTNO			Number				
			7	SCE	ESTSDH			Number				
			-	SCE	ESTMET			Tavt				
			7	SCE				Number				
			7	SCI				Number				
			-	500	STIEIN			Number				
			-	SCO	SMODEL			lext				
			-	SCH	ITAGE			Text				
			-	SCU	JSER1			Text				
			-	SCL	JSER2			Text				
				100					Field Pro	perties		
				Cana								
				Gener	al Lookup	1.1	• *****					
				Field	Size	Int	teger					
				Decin	at Places	Au	ito					
				Input	Mask		10					
				Capti	on						-	
				Defau	ult Value						This property cannot be r	nodified in linked
				Valida	ation Rule						tables.	Iounico III III I
				Valida	ation Text	NZ						
				Index	rea	NC) 0					
				Smart	t Tags		, 					
				Text A	dign	Ge	eneral					
-	50	2 11 6 10 10										
Design vie	ew. ro =	Switch pan	es. Fl = He	elp.							NUN	n Lock
1		3 4	🧊 🚺	X							EN 🔺 🔯 🛱	11:44 a.m. 13/12/2011

Figure 20: DBASE file rendered in Microsoft Access 2007 showing table structure

Some of the database files that were tested included structural components such as field restrictions and links. In Figures 19 and 20 changes can be seen in the field size restrictions and decimal place restrictions. Both of these have been removed in the Microsoft Access 2007 rendering but appear in the DBASE IV rendering.

- File Edit View Incent Form	Microsoft Word - CERTIFY.DOC	▼ ↓
		Ŧ
Normal NewCenturySchlb		
L	······································	<u>1 · · · · 7</u>
	Certificate	
5 		
· · · ·	This is to certify that	
	has successfully attained a high level of skill in	
• • • •		
2 	on.	
· · · · · · · · · · · · · · · · · · ·		
- - - - - - - - -	New Zealand Forest Research Institute	+ 1 1
	Information 16 chinology Urotab	•
Page 1 Sec 1 1/1 At	Ln Col 1:31 PM REC MAK EXT OVA WPH	

MICROSOFT WORD FOR WINDOWS VERSION 2 DOCUMENT FILE

Figure 21: Word for Windows version 2.0 file rendered in Word 6.0c for Windows 3.x

CERTIFY	DOC - LibreOffice Writ	er						
<u>F</u> ile <u>E</u> dit	<u>V</u> iew <u>I</u> nsert F <u>o</u> rmat	T <u>a</u> ble <u>T</u> ools <u>V</u>	<u>V</u> indow <u>H</u> elp					×
B • E			s) X I 🖸 • 🇯	1 (4) + A + 1 🔊 🔳 •	· 🗗 🌚 🔶 🖬 🗃	¶ 🔍 🞯 💂 🗄 🖪	nd 💽 🖓 🗘	
🗄 De	fault 💌	Times New Rom	an 💌 12 💌		I I II II II I	🗉 🗨 🖻 🕍 🕶	s • 🗉 • 🔒	
L		· 6 · 4 · 2	· 🛛 2 · 4 · 6 · 8 · 1	24 · 22 · 20 · 18 ₋ 16 · 14 · 20 · 22	26 · 28 · 30 · 32 · 34 · 36 · 3	38 40 • 42 • 44 • 46 •		*
- 4				Contifi	nato			
-				Ceruin	cate			
-								
			8.	-	-	- 18		
2								
4				I his is to certify that	E.			
-								E
-			h	as successfully attained a high le	vel of skill in			
-								
8				on				
12								
14								
- 0								
-								
- 18								
20								
22			÷					
. 4								-
-				New Zealand Forest Research	h Institute			
- 26				Information 1 echnology	Group			
Page1/1	Default	E	nglish (New Zealand)	INSRT STD	Table1:A1	B 66 £2) - • 	⊕ 75%
1		1 A	2		ELS K	EN	 Image: 11 13/ 	:16 a.m. /12/2011

Figure 22: Word for Windows version 2 file rendered in LibreOffice Writer version 3.3.0

DROID identified the file in this example as a Microsoft Word for Windows version 2 file. When rendered in the nearest version that was available, Microsoft Word version 6.0c the word "certificate" in the document was displayed using the "NewCenturySchlbk" font. This was changed when the file was rendered in LibreOffice Writer 3.3.0 when "certificate" was rendered using the Times New Roman font. This may be a significant issue in certain cases. Font changes have been shown to be very problematic in cases where the font can no longer be found and substitutes are inadequate¹⁹.

In addition to the font change the New Zealand Forest Research Institute logo was removed in the LibreOffice Writer 3.3.0 rendering of the document.

¹⁹ See: "Born Broken: Fonts and Information Loss in Legacy Digital Documents". Brown, Geoffrey; Woods, Kam. 2009. <u>http://escholarship.org/uc/item/53z897zb</u> (accessed 13/12/11)

WORDPERFECT FOR WINDOWS 5.1 DOCUMENT FILE EXAMPLE 1

ile <u>E</u> dit <u>V</u> iew <u>L</u> ayout Tools	Font Graphics Macro Window Help	Samed
VEW ZEALAND MIDDAD N	11 /1051)	
ALW ZEADAND MIKKOK N		
Produced by		
New Zealand National Film U	hit	
MUTTON BIRDERS		
An unusual expedition to outlying islets	at the southern end of New	
Zealand to collect the young of the soo	ty shearwater, commonly known	
as mutton-onus.		
ROUGHRIDERS		
New Zealanders staging an American-s	tyle rodeo, full of spills and	
laugns.		
MAORI ART TREASURES		
A famous collection of Maori art treas	res, made by the late	NEW ZEALAND MIRROR No
IVIr.W. Uldman, of London, has been be Government Old-time weapons tools	ought bt the New Zealand and ornaments made before the	
coming of the white man.	and childheads induc ocidic hic	(1951)
		2888-027 88-07904-0890-880
Film Censor's Certificate No. U2394 (15 Jun	(1541)	
Black & White	10 Minutes	

Figure 23: WordPerfect for Windows 5.1 file rendered in Word Perfect for Windows 5.2

- Previous Pa	Page - View the previous page	- +
<u> </u>	<u>Y</u> iew Insert Fo <u>r</u> mat T <u>a</u> ble <u>G</u> raphics Tools <u>W</u> indow <u>H</u> elp	\$
🗅 🔁 🞜 📇	3 米科は、「「」、「」、「」、「「」、「」、「」、「」、「」、「」、「」、「」、「」、「	
CG Times	▼ 7.2 pt = Styles → Left → 1.0 → Table → Columns → 100% →	
<u>CG Times</u>	************************************	•
Table_A Cell A1	No Printer Select 13 December 2011 1:48PM Pg 1 Ln 3.7" Pos 1.8"	

Figure 24: WordPerfect for Windows 5.1 file rendered in WordPerfect for Windows 6.1



Figure 25: WordPerfect for Windows 5.1 file rendered in Microsoft Office 2007

The WordPerfect for Windows 5.1 file in this example contains a cover sheet for a VHS video tape. Figures 23, 24 and 25 illustrate changes in rendering of the file between different versions of WordPerfect for Windows 3.11 and Microsoft Word 2007.

The fonts change in the Microsoft Word 2007 rendering, two "{PRIVATE}" labels are added to the heading in the top left of the document and the text identifying the VHS as holding a Black and White and 10 minute video clip move their position.

The positioning of the text on the "spine" is also changed between the different renderings. Although not visible, the WordPerfect 5.2 for Windows rendering (Figure 23) does allow for the text on the spine to be displayed via a different interface once the space on the screen is clicked on.



Figure 26: WordPerfect file rendered in WordPerfect 5.2 for Windows



Figure 27: WordPerfect file rendered in Microsoft Office 2007

In the example illustrated in Figures 26 and 27 a file containing a list of films is rendered in the "control" WordPerfect for Windows 5.2 environment and in Microsoft Word 2007. In the Microsoft Word 2007 rendering content is added, the fonts are changed and the formatting is significantly altered.



COREL WORDPERFECT VERSION 7 DOCUMENT FILE

Figure 28: Corel WordPerfect version 7 file rendered in Corel WordPerfect version 7



Figure 29: Corel WordPerfect version 7 file rendered in LibreOffice Writer 3.3.0



Figure 30: Corel WordPerfect version 7 file rendered in Microsoft Office 2007



Figure 31: Corel WordPerfect version 7 file rendered in Corel WordPerfect X5

Figures 28-31 illustrate the rendering of a file containing a report on tree growth. The report contains a number of diagrams including the one illustrated in these examples that identifies properties of a tree at different stages in its lifecycle.

The "control" Corel WordPerfect version 7 rendering of the file and the diagram is virtually identical to the Corel WordPerfect X5 rendering of the file and diagram. The LibreOffice Writer 3.3.0 and Microsoft Word 2007 renderings of the file and diagram are significantly altered. In the LibreOffice Writer 3.3.0 case almost all of the information in the diagram is lost and in the Microsoft Word 2007 case some is lost and some is altered in a way that may affect the meaning of the diagram.

WORDSTAR DOCUMENT FILE



Figure 32: WordStar file rendered in WordStar version 7 for MS-DOS



Figure 33: WordStar file rendered in Microsoft Word 5.5 for MS-DOS

LibreOffice Writer
ert F <u>o</u> rmat T <u>a</u> ble <u>T</u> ools <u>Wi</u> ndow <u>H</u> elp
🖻 🔽 🔚 🖓 I 🖑 🐨 I 🔏 🗇 🗋 • 🏄 I 🄄 • 🔗 • 🚳 📼 • 🜠 I 🖗 🔶 📼 🐿 🦷
1 Text 💽 Times New Roman 💽 10 💽 🙈 🔌 📥 📰 🗄 🗐 🗐 🗐 🏹 🖺 📰 📰 🖛 🕨 📲
· 4 · 2 · X · 2 · 4 · 6 · 8 · 10 · 12 · 14 · 16 · 18 · 20 · 22 · 24 · 26 · 28 · 30 · 32 · 34 · 36 · 38
TAPER EQUATIONS
PREDICTION OF DIAMETER(d) AT ANY LENGTH(L) FROM TIP OF TREE IS :
d = 100 *SqRoot((4 V/(Pi*H)))
(B1(L/H) + B2(L/H) + B3(L/H) + B4(L/H) + B5(L/H) + B6(L/H))
PREDICTION OF VOLUME(v) at A LENGTH(L) FROM TIP OF TREE IS :
$\frac{2}{\chi = (\sqrt{H})^{2} (B1(L/2H) + B2(L/3H) + B3(L/4H) + B4(L/5H) + B5(L/6H)}$
_EQUATION CODES ARE TINN FOR INSIDE BARK function no nn.
COLUMN NUMBER_DESCRIPTION
2- 5EQUATION CODE
7-31 Description 32-39 B1 * 100,000) Normally the first column of 40-47 B2 * 100.000) each of these will be blank

Figure 34: WordStar file rendered in LibreOffice Writer 3.3.0

A set of equations was included in the WordStar file illustrated in Figures 32, 33 and 34. When rendered in all other tested environments including Microsoft Word 5.5 for MS-DOS (Figure 33) and LibreOffice Writer 3.3.0 (Figure 34) the meaning or purpose of the numbers above the equation in the examples was not apparent. When rendered in the control WordStar version 7 environment it becomes clear that the numbers above the equation in the examples are intended to be interpreted as exponents (or "powers" e.g "x²"). This meaning was presumably captured in the control environment by the creator utilizing the font spacing and text spacing functionality to position the exponents above the relevant positions of the equation on the line below.

APPENDIX 3. TIME REQUIRED TO TEST THE RENDERING OF OBJECTS

Testing approaches is a difficult challenge for digital preservation practitioners. It is a potentially very costly aspect of digital preservation.

The table below outlines how long it would take for a variety of different sample sizes and proportions of objects to be tested assuming a similar 9 minutes to test the results of rendering or migration of each object using a single tool.

Table 10: Time taken to test x percentage of objects (hours)

					Percentage	e tested (%)			
Number of Objects	0.5	1	2	5	10	25	50	75	100
100	0.075	0.15	0.3	0.75	1.5	3.75	7.5	11.25	15
1,000	0.75	1.5	3	7.5	15	37.5	75	112.5	150
5,000	3.75	7.5	15	37.5	75	187.5	375	562.5	750
10,000	7.5	15	30	75	150	375	750	1125	1500
25,000	18.75	37.5	75	187.5	375	937.5	1875	2812.5	3750
50,000	37.5	75	150	375	750	1875	3750	5625	7500
100,000	75	150	300	750	1500	3750	7500	11250	15000
250,000	187.5	375	750	1875	3750	9375	18750	28125	37500
500,000	375	750	1500	3750	7500	18750	37500	56250	75000
1,000,000	750	1500	3000	7500	15000	37500	75000	112500	150000
2,000,000	1500	3000	6000	15000	30000	75000	150000	225000	300000
5,000,000	3750	7500	15000	37500	75000	187500	375000	562500	750000
10,000,000	7500	15000	30000	75000	150000	375000	750000	1125000	1500000

Table of time taken to test x percentage of objects (hours)

APPENDIX 4. QUESTIONS IN DOCUMENTATION QUESTIONNAIRE

The full digital questionnaire that was used in this research will be released alongside this report in an open format. The questions included in the questionnaire are listed below. The question routing is not described below but is included in the digital questionnaire file.

Table 11: Questions in documentation questionnaire

QuestionID	Question Text
829	Digital object type:
830	Digital Object ID (if available)
831	Computer File ID From DROID
832	Test Application ID (repeated below)
833	Will the object render?/Will the file open at all in the application?
834	Are there macros or scripts in the digital object?
835	Are there any links in the file to other files?
836	Can the Macros or Scripts be executed?
837	Are the links to external files still working?
838	Are there any editing restrictions on the object?
839	Have the restrictions been maintained?
840	What type of rendering is being observed in this test?
841	Does the object contain an edit history?
842	Has the edit history been maintained?
843	Is there metadata embedded in the file such as the author's name, date saved,
	amount of time spent authoring, etc?
844	Has the embedded metadata been maintained?
845	Are any/all fonts being fully and accurately rendered?
846	Has the text formatting been maintained? e.g. bold, italic, underline, superscript, sub
847	Is there text formatting included in the object e.g. hold italics underline strike-
011	through, subscript or superscript?
848	Does the object have text of any colour other than black?
849	Has the text colour been maintained?
850	Does the object include highlighted text?
851	Has the highlighted text been maintained?
852	Have the page dimensions been maintained?
853	Has the pagination been maintained?
854	Has the position on screen of content been maintained?
855	Has the position of content on the page been maintained?
856	Has line spacing been maintained?
857	Have the new-lines been correctly placed?
858	Have page and section breaks been maintained?
859	Has the orientation of objects/text been maintained?
860	Has the justification of text been maintained?

861	Has any extra/additional information/data been added to the object that is
	observable by the user?
862	Does the object contain images?
863	Has the image orientation and position been maintained?
864	Has the image size been maintained?
865	Have the colours of the image been maintained?
866	Has the resolution of the image been maintained?
867	Does the object include custom views?
868	Have custom views been maintained?
869	Does the object include custom shapes?
870	Have custom shapes been maintained?
871	Does the object include hidden content?
872	Has the hidden content been maintained?
873	Does the object include watermarks?
874	Have the watermarks been maintained?
875	Does the object include custom character sets?
876	Have the custom character sets been maintained?
877	Does the object include any custom languages or language interfaces?
878	Have the custom languages or language interfaces been maintained?
879	Has the number of words reported by the software been maintained?
880	Has the actual number of words in the document been maintained?
881	Does the document have footnotes or endnotes?
882	Have the footnotes or endnotes been maintained?
883	Does the document have an embedded object that adds the current-date to the object?
884	Has the embedded date been maintained? (please comment)
885	Does the document have internal links within it?
886	Have the internal links been maintained?
887	Does the document include lists or bullet points?
888	Have the lists or bullet points been maintained?
889	Have the list or bullet point symbols been maintained?
891	Have the tables been maintained?
892	Has the table formatting/layout been maintained?
893	Are there borders within the document?
894	Have the borders been maintained?
895	Are there citations in the document?
896	Have the citations been maintained?
897	Are there mail-merge settings applied in the document?
898	Have the mail-merge settings been maintained?
899	Does the document include comments?
900	Have the comments been maintained?
901	Are there formulae in the object?
902	Have the formulae been maintained?

903	Has the notation language of the formulae been maintained?
904	Has the way rounding is calculated been maintained?
905	Has the number of decimal places displayed been maintained?
906	Does the object have any internal links?
907	Have the internal links been maintained?
908	Does the object have an embedded object that adds the current-date to the object?
909	Has the embedded date been maintained? (please comment)
910	Does the object have coloured cells?
911	Have the cell colours been maintained?
912	Does the object include cells with borders?
913	Have the cell borders been maintained?
914	Does the object include any conditional formatting?
915	Has the conditional formatting been maintained?
916	Has the column order been maintained?
917	Has the row order been maintained?
918	Have functions been maintained? E.g. standard deviation?
919	Does the object include pivot tables?
920	Have the pivot tables been maintained?
921	Does the object include hidden rows or columns?
922	Have the hidden rows or columns been maintained?
923	Does the object include named cells?
924	Have the cell names been maintained?
925	Does the object include named ranges?
926	Have the name ranges been maintained?
927	Have cell types been maintained? E.g. number, text or date
928	Does the object include any applied filters?
929	Have the applied filters been maintained?
930	Does the object include links to other data sources?
931	Have the links to other data sources been maintained?
932	Does the object include multiple worksheets?
933	Have all of the worksheets been maintained?
935	Has the embedded date been maintained? (please comment)
936	Has the shading or colours been maintained?
937	Has the layout been maintained?
938	Does the graph include a title?
939	Has the tile been maintained?
940	Does the graph include labels on axes or data points?
941	Have the labels been maintained?
942	Have the proportions and/or ratios of the axes been maintained?
943	Does the graph include the ability to view the data source(s)?
944	Has the ability to view the data source(s) been maintained?
945	Does the Presentation include animated (or other) slide transitions?
946	Have the slide transitions been maintained?

947	Does the presentation include audio or video?
948	Is the audio or video still renderable?
949	Has the interface/presentation mode of the slides been maintained, for example the click to do x or wait 2 seconds before x?
950	Does the presentation have an embedded function that adds the current-date to the object?
951	Has the embedded date been maintained? (please comment)
952	Does the database have a custom front-end, interface or form(s)?
953	Has the ability to render and interact with the custom front-end been maintained?
954	Does the database include saved queries?
955	Have the queries been maintained?
956	Has the internal structure been maintained (e.g. primary keys, links between tables etc).
957	Does the database include links to other data sources?
958	Have the links to other data sources been maintained?
959	Does the database include any custom views?
960	Have the custom views been maintained?
961	Does the database have an embedded function that adds the current-date to the object?
962	Has the embedded date been maintained? (please comment)
963	Has all useful functionality in the object been maintained?
964	Are there any other changes to the object that have not been identified in other questions?
965	Any other comments or notes?
966	Please list the file names and/or locations of any screenshots you have made to document this test

APPENDIX 5. APPLICATIONS USED IN TESTING

Table 12: Applications used in testing

AppID	appName	version	testOperatingSystem
34	Microsoft Excel 97 SR-1	97 SR-1	Windows 98 SE (4.10.2222 A)
33	Microsoft Word 97 SR-1	97 SR-1	Windows 98 SE (4.10.2222 A)
74	Paradox 7	Version 7.0	Windows 98 SE (4.10.2222 A)
18	Microsoft Access for Windows 95	Version 7.00	Windows 95
20	Microsoft Excel for Windows 95	Version 7.0	Windows 95
26	Microsoft Word for Windows 95	Version 7.0	Windows 95
29	Microsoft Works Word Processor Version 4.0 for Windows 95	4	Windows 95
8	Corel WordPerfect	Version 6.1 for Windows	Windows 3.1
10	Freelance Graphics for Windows	Release 2.1	Windows 3.1
19	Microsoft Access Version 2.00	Version 2.00	Windows 3.1
21	Microsoft Excel Version 5.0c	Version 5.0c	Windows 3.1
27	Microsoft Word Version 6.0c	Version 6.0c	Windows 3.1
30	Quttro Pro 6.02	Version 6.02	Windows 3.1
31	WordPerfect for Windows	Version 5.2	Windows 3.1
6	Ashton Tate DBASE IV	Verison IV	MS-DOS 6.22
9	Framework II	11	MS-DOS 6.22
----	--	---------------------------	---
63	Microsoft Word	5.5	MS-DOS 6.22
62	WordPerfect 5.1 +	5.1	MS-DOS 6.22
64	WordStar for DOS North American Version 7.0 Rev. A	7.0 Rev. A	MS-DOS 6.22
59	Microsoft Office Access 2003 (11.8328.8329) SP3 Part of Microsoft Office Professional Edition 2003	11.8328.8329	Microsoft Windows XP [Version 5.1.2600]
56	Microsoft Office Word 2003 (11.8328.8329) SP3 Part of Microsoft Office Professional Edition 2003	11.8328.8329	Microsoft Windows XP [Version 5.1.2600]
48	Corel Presentations X5	15.0.0.357	Microsoft Windows 7 Professional Version 6.1.7600
46	Corel Quattro Pro X5	15.0.0.357	Microsoft Windows 7 Professional Version 6.1.7600
47	Corel WordPerfect X5	15.0.0.357	Microsoft Windows 7 Professional Version 6.1.7600
53	LibreOffice Base 3.3.0 OOO330m19 (Build:6)	3.3.0 OOO330m19 (Build:6)	Microsoft Windows 7 Professional Version 6.1.7600
50	LibreOffice Calc 3.3.0 OOO330m19 (Build:6)	3.3.0 OOO330m19 (Build:6)	Microsoft Windows 7 Professional Version 6.1.7600

54	LibreOffice Draw 3.3.0 OOO330m19 (Build:6)	3.3.0 OOO330m19 (Build:6)	Microsoft Windows 7 Professional Version 6.1.7600
52	LibreOffice Impress 3.3.0 OOO330m19 (Build:6)	3.3.0 OOO330m19 (Build:6)	Microsoft Windows 7 Professional Version 6.1.7600
51	LibreOffice Writer 3.3.0 OOO330m19 (Build:6)	3.3.0 OOO330m19 (Build:6)	Microsoft Windows 7 Professional Version 6.1.7600
45	Microsoft Office Access 2007 (12.0.4518.1014) MSO (12.0.4518.1014) Part of Microsoft Office Professional 2007	12.0.4158.1014	Microsoft Windows 7 Professional Version 6.1.7600
41	Microsoft Office Excel 2007 (12.0.4518.1014) MSO (12.0.4518.1014) Part of Microsoft Office Professional 2007	12.0.4158.1014	Microsoft Windows 7 Professional Version 6.1.7600
43	Microsoft Office PowerPoint 2007 (12.0.4518.1014) MSO (12.0.4518.1014) Part of Microsoft Office Professional 2007	12.0.4518.1014	Microsoft Windows 7 Professional Version 6.1.7600
44	Microsoft Office Publisher 2007 (12.0.4518.1014) MSO (12.0.4518.1014) Part of Microsoft Office Professional 2007	12.0.4158.1014	Microsoft Windows 7 Professional Version 6.1.7600
42	Microsoft Office Word 2007 (12.0.6514.5001) MSO (12.0.4518.1014) Part of Microsoft Office Professional 2007	12.0.6514.5001	Microsoft Windows 7 Professional Version 6.1.7600
75	Microsoft Word 2000	9.0.3821 SR-1	Microsoft Windows 2000 5.00.2195 Service Pack 4