*Title:* **PAWDOC Preservation MAINTENANCE PLAN**

*Author:* Paul Wilson

*Date:* 28th May 2018

**NB.** v1.1 was produced to reflect revised numbers in Section 2 and the inclusion of section 3.13, as a result of the writing of the paper describing the PawdocDP project.

**Schedule:** The next Maintenance exercise is scheduled to start on 1st September 2021. To perform the maintenance, go through each of the following 16 steps.

**1. Purpose:** This document specifies the actions to be taken in the future to enable the contents of the PAWDOC collection (both digital and physical) to continue to be viewed and used to a level of quality and utility not less than was possible in 2018.

**2. Previous preservation actions taken:** At the point at which this document was written, the following preservation actions had been taken:

* **DMS removed:** The use of the Fish Document Management System (DMS) was discontinued because its supplier/owner had no plans to develop it for future platforms and it was too complicated and expensive to maintain. This was achieved by exporting the files out of Fish into Windows folders. Extensive checking of the export logs indicated the export had been a success (full details are available in the file 'PAW-DOC-9189-02 Fish Export checks v1.0 - 03Mar2018, 06-03-2018.xlsx'). This was further verified in the subsequent preservation activities in which the contents of over 1,600 of the 17,200 folders were reviewed without identifying any problems due to the export.
* **Zip files checked:** All 93 zip files in the collection were checked (full details are available in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 Droid Analysis for PawdocDP Scoping Work v3.2, 28May2018'). Where possible, action was taken to improve their future accessibility. This included:
  + 2 Lotus Notes files were converted to 2 PDF files and 1 XLSX file.
  + 15 Powerpoint files and 9 Visio files all within one larger zip were converted to 15 Powerpoint 1997-2003 files and 9 PDF files respectively and re-zipped.
  + 22 zips containing 199 eml files were converted to 22 PDF files.
  + 2 zips containing 66 GIFs were converted to 2 PDF files.
  + 2 zips containing 28 html files were converted to 2 PDF files.
  + 6 zips are recorded in sections 3.1 - 3.6 of this document as possibly requiring work in the future.
  + 18 zips containing a variety of file types were converted to 68 separate Word, Excel, Powerpoint, PDF, JPG, CFG, ID, MOV, DLL and EXE files.
* **File types analysed using DROID**: An analysis of all the Collection's file types was carried out using the DROID tool (full details are available in the file spreadsheet 'PAW-DOC-9189-03 Droid Analysis for PawdocDP Scoping Work v3.2, 28May2018'). This identified some very old formats which it was decided to upgrade, and, as a result, the following files were converted:
  + 10 Microsoft Visio (generic) files and 2 Visio Drawing 2000 files were converted to 12 PDF files.
  + 1 WordPerfect for MS-DOS/Windows Document 5.1 file was converted to 1 PDF file.
  + 20 Microsoft Word for Macintosh Document 3.0 files were converted to 20 DOCX files.
  + 100 Microsoft Word for Macintosh Document 4.0 files were converted to 78 PDF and 22 DOCX files.
  + 1 Microsoft Word for Macintosh Document 5.0 file was converted to 1 DOCX file.
  + 30 Microsoft Word for Windows Document 1.0 files were converted to 25 PDF and 5 DOCX files
  + 436 Microsoft Word for Windows Document 2.0 files were converted to 384 PDF and 52 DOCX files
  + 3 Microsoft Word Document Template 97-2003 files were converted to 2 PDF and 1 DOCM files
  + 4 Microsoft Excel 3.0 Worksheet files were converted to 4 XLSX files.
  + 17 Microsoft Excel 4.0 Worksheet files were converted to 17 XLSX files.
  + 56 Powerpoint for Macintosh 4.0 files were converted to 56 PPTX and 56 PDF files
  + 2 Microsoft Powerpoint Presentation 4.0 files were converted to 2 PPTX and 2 PDF files
  + 4 MIME Email 1.0 files were converted to 1 Word 1997-2003 file and 3 PDF files
  + 7 Rich Text Format 1.0-1.4 files were converted to 7 PDF files
  + 6 Rich Text Format 1.5-1.6 files were converted to 6 PDF files
  + 1 Rich Text Format 1.7 file was converted to 1 PDF file
  + 1 Rich Text Format 1.9 file was converted to 1 PDF file
  + 6 files are recorded in sections 3.7 and 3.8 of this document as possibly requiring work in the future
* **Non-opening files investigated:** A previous check had identified 220 files in the collection which would not open. Each of these was investigated to identify a solution (full details are available in the file ' PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). Action taken included the following:
  + 2 Windows Help Files were converted to 2 CHM files.
  + 1 TIF file was converted to 1 PDF file.
  + 1 Lotus 1-2-3 Worksheet 1.0 file was converted to 1 XLSX file.
  + 2 iThink files were converted to 2 PDF files.
  + 3 Microsoft Access Database 2.0 files were converted to 3 XLSX files.
  + 1 Microsoft Word for Macintosh Document 4.0 was converted to 1 DOCX file.
  + 2 Microsoft Word (Generic) 6.0-2003 files were converted to 3 DOCX files.
  + 33 Powerpoint for Macintosh 4.0 files were converted to 33 PPTX and 33 PDF files.
  + 58 Microsoft Powerpoint Presentation 4.0 files were converted to 58 PPTX files.
  + 58 Microsoft Powerpoint Presentation 97-2003 files were converted to 58 PPTX files.
  + 20 files still cannot be accessed and all have been recorded in sections 3.9 - 3.13 as requiring further work in the future.
* **Files converted to PDF to re-order contents:** 881 multi-page TIF files were converted to PDF in the course of reordering them after having been originally digitised on a single-sided scanner by scanning one side of the page pile first and then the reverse sides (full details are available in the file 'PAW-DOC-9189-03 - Double Unsorted Work List v1.0, 25Apr2018').
* **Backup disks copied to the laptop:** The contents of 125 backup disks were copied to the APAWBACK folder; and the accessibility of the disks in the collection was reviewed.
* **Backup/DR arrangements revised:** New backup and disaster recovery arrangements were documented in the PAWDOC User Guide (see PAW-DOC-9189-04).

**Note** that all files that were converted or changed in any way in the above preservation activities kept the same file title as the original but had 'UPDATED' and the date of the conversion/change inserted into their titles. The original versions were retained but had 'ORIGINAL' inserted into their file titles. **Both UPDATED and ORIGINAL files reside side by side inside the collection's folders.**

**3. Possible future preservation issues:** In the course of the above preservation actions being taken, the following issues were identified as possibly requiring preservation work in the future:

**3.1 PAW-DOC-4006-01**

This contains the zip file 'Caseware demo files, 26-02-2005, A' which in turn contains two Caseware demo files - CMD--ZIP.EXE and CWDEMO.BAT. These would not run under Windows 7 and no further action was taken (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018' ). There are also two Caseware brochures in the same Fish folder so the fact that this doesn't work isn't disastrous, however further work should be undertaken to see if it is possible to get these files to run. PAW-DOC-5928-01 contains ten data files for the Casewise process modelling software (see section 3.12). However, the origins or current owner of this software, and its possible relationship to Caseware, could not be identified. Further efforts to establish if there is a relationship between Casewise and Caseware should be considered for the next set of digital preservation work.

**3.2 PAW-DOC-4052-01**

This contains the zip file ' Isovu presentation provided on floppy disk, 09-04-2005, 15' which in turn contains 54 files in all - 40 .pic files, 6 .sh~ files, 2 .txt files, 1 .bat file, 1 .str file, 3 .exe files, and 1 .com file. These would not run under Windows 7 and no further action was taken (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018' ). There is an ISOVU brochure in the same Fish folder so the fact that this doesn't work isn't disastrous. However, when the platform is moved to a new operating system, it is probably the right time to consider what to do, if anything, with these files.

**3.3 PAW-CD-008-01**

This contains the zip file CDI 2000 which contains the contents of a physical CD which is still in the collection. The CD contains hundreds of files, however to get it to work it is necessary to run the WIN97 installer - which doesn't work under Windows 7. Furthermore its possible that some of the Word and PowerPoint files may be too old to run as well (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018' ). It was too big a challenge to get this to work in the course of the PawdocDP project during 2017-18. Therefore, investigations into getting the contents of this CD to work as originally intended, need to be scheduled into the next set of Preservation Maintenance work that takes place.

**3.4 PAW-CD-002-01**

This contains the zip file 'An introduction to e-Business, Oct1999, 18-05-2016, 55' which contains the contents of a physical CD which is still in the collection. The CD contains hundreds of files, however it won't open under Windows 7 (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018' ). It was too big a challenge to get this to work in the course of the PawdocDP project during 2017-18. Therefore, investigations into getting the contents of this CD to work as originally intended, need to be scheduled into the next set of Preservation Maintenance work that takes place.

**3.5 PAW-CD-004-01**

This contains the zip file 'PC Magazine CD April 2001, 19-05-2016, 59' which contains the contents of a physical CD which is still in the collection. The CD contains many shareware programs. McAfee identifed a trojan when these files were zipped in 2016. The trojan and several associated files were removed from the zip, but they will still be on the CD. It is not known what form the trojan takes (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 8May20018' ). It was too big a challenge to go through the whole of this disk in the course of the PawdocDP project during 2017-18. Therefore, consideration of what needs to be done with the contents of this zip file, needs to be scheduled into the next set of Preservation Maintenance work that takes place.

**3.6 PAW-DOC-3889-01**

This contains the zip file 'Nautilus CD, 25-05-2016, 66' which contains the contents of a physical CD which is still in the collection. The CD was a 1991 promo disk for a disk which was going to be issued every month containing a variety of multimedia resources, software, games, and technical advice and news. Most of the base files seem to be either text files or GIFs. It won't open because the main screen is a file of.tbk (an application called Toolbook) which won't open with Firefox or Acrobat. It opens in Notepad but without the links. The READMEs suggest that setup.exe should be run first but that won't run under Windows 7 (further details are in the 'Zip Files' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018'). It was too big a challenge to get this to work in the course of the PawdocDP project during 2017-18. Therefore, investigations into getting the contents of this CD to work as originally intended, need to be scheduled into the next set of Preservation Maintenance work that takes place.

**3.7 PAW-DOC-3906-104**

This contains the two files 'Beneficial loan calculator program' and 'dll file to work with the executable'. These would not run under the present version of Windows 7 (though the lack of an XX entry in the Index indicates that it DID run on the laptop before the crash and rebuild in Nov2017); and no further action was taken (further details are in the 'File Type Analysis' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018'). However, when the platform is moved to a new operating system, it is probably the right time to consider what to do with these files.

**3.8 PAW-DOC-4581-01, PAW-DOC-4389-01, PAW-DOC-4439-01**

None of these four files (PAW-DOC-4581-01 (Fish file ref 1944.doc and 1945.doc), PAW-DOC-4389-01 (Fish file ref 194B.doc) + PAW-DOC-4439-01 (Fish file ref 1F2C.doc) can be saved in DOCX format. They all give the message 'There is not enough memory to display or print the picture' when paging through the document and they all have empty spaces for some diagrams (further details are in the 'Action on File Types' worksheet of the spreadsheet 'PAW-DOC-9189-03 - Droid Analysis for PawdocDP Scoping Work v3.2, 28May20018'). This should be investigated again when a new platform is acquired and if the problem persists it may be best to just save in PDF format with blank spaces for the missing diagrams.

**3.9 PAW/DOC/6311/01, PAW/DOC/6993/01, PAW/DOC/6243/01, PAW/DOC/6207/03**

The internet Zamzar file conversion service was unable to convert the following four PowerPoint files into a version that would display in Powerpoint 2007: PAW/DOC/6311/01 (3D.ppt), PAW/DOC/6993/01 (10B.ppt), PAW/DOC/6243/01 (120.ppt), PAW/DOC/6207/03 (121.ppt). Droid identified the first two as Microsoft PowerPoint for Macintosh (4.0) files, and the second two as Microsoft Powerpoint Presentation, 4.0 files (further details are in the file PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). A service capable of converting these files into readable documents should be sought in the next set of preservation work.

**3.10 PAW-DOC-6036-01**

This contains a Microsoft Project file (Fish file ref - 1.mpp) for which no converter could be found (one source which tried to open it reported " I tried opening the file using Project 98 and got an error message saying there was a problem with the file so either it is corrupt or from a version of Project earlier than 4.x." (further details are in the file PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). One final effort to investigate this file and to try and convert it, should be made in the next set of digital preservation work.

**3.11 PAW-DOC-6612-01**

This contains four Lotus Screencam files (Fish file refs - 6.scm, 7.scm, 8.scm, 9.scm) for which no converter could be found (further details are in the file 'PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). One final effort to try and convert these files should be made in the next set of digital preservation work.

**3.12 PAW-DOC-5928-01**

This contains ten data files for the Casewise process modelling software for which a conversion capability could not be found (further details are in the file 'PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). Further efforts to convert all these files, should be made in the next set of digital preservation work. Note also the possible relationship to Caseware files mentioned in section 3.1.

**3.13 PAW/DOC/6518/42**

This contains a Lotus Organiser v2 file for which a converter has not been found (further details are in the file 'PAW-DOC-9189-03 PAWDOC Files that won't open - Categories and Details, v2.1, 28May2018'). A further effort to convert this file should be made in the next preservation Maintenance exercise.

**3.14 PDF files of A5 brochures**

34 PDF files contain scans of A5 brochures with the pages in incorrect order. Although this is not a Preservation problem as such, it is recorded here because it is an issue remaining from the 2018 preservation exercise. Rectification of the problem will involve inserting a copy of each page into the PDF and cropping out one of the A5 pages on each of the copies; and then reordering all the pages. The items concerned can be identified by searching for 'A5-unsorted' in the Index Movements field. Further details are available in the file ''PAW-DOC-9189-03 - Double Unsorted Work List v1.0, 25Apr2018'.

**3.15 Non-conformance to the PDF-A-1b standard**

PDF files created in the previous digital preservation exercise were not conformant to the PDF-A-1b standard, and the eCopy PDF Pro 'FIX' facility was unable to rectify all of the non-conformances. Consideration needs to be given as to whether it is necessary to undertake work to ensure that all PDF files in the collection comply fully with the PDF-A-1b standard.

**4. How to implement this Maintenance Plan:** To implement this Plan, perform the activities listed in this section below. A description of each activity is provided in sections 5 - 16 of this document. If a more rigorous approach is required, create a Digital Preservation Plan DESCRIPTION and CHART (possibly preceded by a SCOPING document), at the points indicated in the steps below. The activities are to be performed in the order they are listed below:

* specify the preservation period
* review the overall hardware and operating system strategy
* perform an inventory check and identify any work required on the physical items and storage
* catalogue the file types
* then, for each file type:

- try opening a few example files

- define the application and platform strategy for that file type

- define any conversion action to be taken.

If required, create a SCOPING document, and a Digital Preservation Plan DESCRIPTION and CHART at this point.

* make any changes to the Applications and Platforms that have been defined
* perform any work previously specified on the physical items and storage
* for each file type, implement the specified file conversion actions
* create the next version of this maintenance plan
* review and update the backup arrangements and make new backups as necessary.

**5. Specify the preservation period:** Decide when the next Maintenance exercise after this one is to take place. This will determine the minimum period over which the collection is to be preserved by the actions taken in this maintenance exercise. It is recommended that the period should not exceed 5 years.

**6. Review the overall hardware and operating system strategy:** The ability to open and use the collection's digital files is critically dependent on the hardware and operating system platforms upon which the application software runs. Therefore, identify if any changes are planned for these platforms over the period leading up to the next maintenance exercise.

**7. Perform an inventory check and identify any work required on the physical items and storage:** Check that all physical and digital items are in place. While doing so, identify any physical items or storage containers that are in poor condition or inadequate in any way, and specify what action is to be taken to rectify the problems. Try to find any missing items. If they still can't be found, record the fact that they are missing.

**8. Catalogue file types:** Use a software tool such as DROID (which was downloadable for free from the National Archives web site in 2018), or simply use Windows Explorer or equivalent in other operating systems, to identify what file types are present in the collection.

**9. Try opening a few example files:** For each file type, take a few example files and try to open them. Make notes on what happens including what application (and its version number) is used to open the files.

**10. Define the application and platform strategy:** For each file type, decide what application you want to be able to open them in; then what operating system you want the application to run on; and then what hardware you want to use (it is probable that you'll want to use the same operating system and hardware platforms for all the items in the collection - but it is also possible that particular applications may influence the platform decision).

**11. Define conversion action to be taken:** For each file type, define what conversion, if any, is to be performed on the files of that type; and the process and tools to be used to achieve the conversion.

**12. Make any changes to the Applications and Platforms that have been defined:** Before proceeding with any conversion activities, make any changes to the Hardware, Operating System and Applications that are required.

**13. Perform any work previously specified on the physical items and storage:** If any problems with the physical items and storage containers had been identified while undertaking the inventory check, carry out the rectification work that has been specified.

**14. Implement the specified conversion actions on each file:** For each file type perform the conversion actions that have been specified on the files of that type. Test that the converted files open up satisfactorily before deleting any originals. Consider retaining the originals to mitigate against any reduction in quality introduced during major conversions from one format to a completely new format (as opposed to just updating the version of a particular format).

**15. Review and update the next version of this maintenance plan:** Go through this document and update it as necessary with information for the next maintenance exercise. Be sure to change the date in the Schedule section at the beginning of the document and to include the date in your personal calendar. In addition, update the ' Previous preservation actions taken' and the 'Possible future preservation issues: ' sections. Modify the maintenance process steps if required as a result of experience gained in the exercise just completed.

**16. Review and update the backup arrangements:** Review the backup arrangements and check they have been operating correctly. Test out the recovery of files from the different backup mechanisms. Make any changes to the backup arrangements that are needed. When all changes have been made, check that the backup documentation is completely up to date, and then ensure that the latest versions of all the electronic components of the PAWDOC collection are backed up.

**ENDS**

**Changes needed to this template**

Section 1: Change ' **Schedule:** The next Maintenance exercise is scheduled to take place on @@@.' to **Schedule:** The next Maintenance exercise is scheduled to start on @@@.

Section 2: change the suggested text to '- The formats of all @@@ files comprising the digital collection had been reviewed and the following conversions performed to ensure the files concerned remain readable in the future'

Section 2: All files that were converted kept the same file title as the original but had 'UPDATED' and the date of the conversion inserted into their titles. The original versions were retained but had 'ORIGINAL' inserted into their file titles. <*Remove this last sentence if it does not apply.*>

Section 2: the text '- Backup arrangements were due to be reviewed, documented and implemented immediately following the production of this document (this was left to last to ensure that this Maintenance Plan gets included in any backups taken as part of that exercise).' is not correct in this instance because it was done prior to the production of the maintenance plan.

Section 3: Change the introductory sentence to '**3. Possible future preservation issues:** In the course of the above preservation actions being taken, the following issues were identified as possibly requiring preservation work in the future: '

Section 3: Probably better to number these items so they can be referred to in other planning documents.

Section 4: Change the introductory para to "**4. How to implement this Maintenance Plan:** To implement this Plan, perform the activities listed in this section below. A description of each activity is provided in sections 5 - 16 of this document. If a more rigorous approach is required, create a Digital Preservation Plan DESCRIPTION and CHART (possibly preceded by a SCOPING document), at the points indicated in the steps below."

Section 4: Change the para in the middle of the list to " If required, create a SCOPING document, and a Digital Preservation Plan DESCRIPTION and CHART at this point."

Section 4: change the line within the file type sub-list to, 'define the application and platform strategy for that file type'

Section 4: change dashes to bullets for the top level of the list.