*Title:* **PAWDOC** **Preservation Planning Scoping Document**

*Author: Paul Wilson*

*Date: 05Sep2017*

1. **What is the name of the collection and who does it belong to?**

The collectionis called Paul Wilson’s Work Document Collection. It’s short name is PAWDOC and it is currently owned by Paul Wilson.

1. **What are the main contents of the collection?**

The collection consists of most documents that Paul Wilson read and produced in the course of his work from June 1981 to when he retired in 2012; and some professional-type documents after his retirement to the present day. It also includes those documents that he thought were of particular value or note from the start of his working career with Kodak in 1972 up to May 1981; and a few documents from his time as a student at Loughborough University, initially on the Civil Engineering course, and, subsequently, on the Ergonomics course. Government Classified documents, and highly confidential Commercial documents such as Bid documents, were not included in the collection.

1. **Why do you want to keep this collection?**

There are three main reasons for keeping this collection:

First, the collection provides a unique insight into professional working life over the period of transition from manual office work to computer-supported office work. It provides this insight in terms of both the impact on the individual (exemplified by material relating to the owner’s working life), and the impact on organisations (illustrated by material relating to the work of the collection’s owner and his employers in assisting others to implement new computer technology and systems).

Second, the collection contains a large number of insightful materials, for example, case studies of over 50 organisations; full POLDAT (Process, Organisation, Location, Data, Application and Technology) architectures of a number of large organisations; extensive documentation about the development of rationales, methodologies and approaches to introduce new technologies into organisations; major usability engineering initiatives; and documentation regarding the early development of the field of Computer-Supported Cooperative Work.

Third, the collection was extremely hard to assemble, to maintain, and to finally convert to full digital form; and therefore it is not something one would want to discard lightly.

1. **For whom are you keeping it? What are their functionality, technology, and any other requirements? How are you going to test their expectations?**

A long term home is being sought for the collection. However, until a home is found, the collection will be managed by Paul Wilson and subsequently Matt Fox-Wilson. The requirements of each of these parties are described below:

**4.1 Paul Wilson Requirements**

Functionality: The ability to look something up in the index and then to find and access all the electronic documents associated with the selected index entry.

Technology: Windows Laptop

Other Requirements: Minimum maintenance, backup and digital preservation effort at no extra cost over and above the purchase of the laptop and the standard Microsoft Office toolset.

**4.2 Matt Wilson Requirements**

Functionality: The ability to demonstrate the main components of the system and the way the index can be used to find and access all the electronic documents associated with a selected index entry.

Technology: Apple System

Other Requirements: Minimum maintenance, backup and digital preservation effort at no extra cost over and above the purchase of the laptop and the standard Apple Office toolset.

**4.3 Eventual Owner once a home has been found for the collection, Requirements**

The eventual Owner is likely to be an individual historian of some sort, representing a group of historians; or an organisation with an interest in modern history.

Functionality: The ability to look something up in the index and then to find and access all the electronic documents associated with the selected index entry. The ability to provide access to the collection to selected others under controlled conditions.

Technology: Laptop system

Other Requirements: Minimum maintenance, backup and digital preservation effort at no extra cost over and above the purchase of the laptop and a standard Office toolset.

1. **What are the main digital components in the collection?**

| **Component Name** | **Contents** | **Technology** | **Physical Equivalents** |
| --- | --- | --- | --- |
| 1. Filemaker software | Filemaker is the database software package which is used to record the indexes to the collection. | The Filemaker software package is supplied by Filemaker Inc – an Apple subsidiary. Version Pro 15 is currently being used (though this still produces files with the extension ‘fmp12’). The Filemaker software resides in the directory c:\Filemaker on an Acer Aspire 4830T Timeline X laptop – referred to subsequently as ‘The Collection Laptop’ or TCL. |  |
| 1. Set Index | This is a record of all the subsets of documents in the collection’s Index e.g. PAW/DOC, PAW/CD, PAW/CHI, RMC/DOC, NCC/CJA. | The information is stored in a Filemaker database called ‘Pawset.fmp12’ which resides on ‘The Collection Laptop’ or TCL. | None |
| 1. Document Index | This is the main Index to the collection. It is a key file and needs regular backup. | The information is stored in a Filemaker database called ‘Pawdoc.fmp12’ which resides on TCL. | None |
| 1. Fish software | Fish is the document management software package which is used to store the electronic files associated with the Document Index entries. | The FISH software package is supplied by m-hance. Currently Version 5.5.5 is being used. Fish was originally supplied by DDS, then by Ringwood Software, then by Maxima and now by the current supplier m-hance. The Fish software resides on TCL in the directory c:\MAXIMA. | None |
| 1. Fish file name control files | These files control the character combinations used to name the latest item of a particular file type stored in the document management system. The information enables the document management system to allocate a unique file name to the next document of a particular file type that it is asked to store. | The files are stored in a directory called STORAGE with the path c:\MAXIMA\FiShServer and which contains a number of different files – each file recording the next character combination to be used to name a particular application file type. For example, the file HIGHEST.JPG records the next character combination used to name the next JPG document to be stored in Fish. | None |
| 1. Fish Configuration File | A new version with a new date is saved each time FISH is used, so it is backed up regularly. | The file is called ‘PCLIPWIN.INI’ and it is stored in the ‘FiShServer’ directory with the path c:\MAXIMA. | None |
| 1. Fish Databases | These two database files store information about the documents that the Document Management System is managing. One is the main database and the other is a log file. | These are Microsoft SQL files called ‘fish\_Data.MDF’ and ‘fish\_Log.LDF’, and they are stored in the ‘FiShDatabaseFiles’ directory with the path c:\MAXIMA. | None |
| 1. SQL Server software | This is the database software used to record control information generated by the FISH document management system. | This is the SQL Server database package from Microsoft – MS SQL 2008 R2. Currently Version 2009.100.1617.0 is being used, for which a critical update was issued in June 2014 but which has not been installed because there is uncertainty about its effect on the performance of Fish. The SQL software resides in the Microsoft SQL Server directory with the pat c:\Program Files | None |
| 1. Fish Documents | These are the actual electronic documents comprising the PAWDOC collection and which are being managed by Fish. | Documents stored in Fish are given unique filenames of the type CE6.pdf and 1FA.jpg and 25B.ppt. i.e. a combination of letters and numbers – each one unique within a particular file type. File types stored within the collection include: TIF, PDF, Word, PowerPoint, Excel, HTML, Help, Zip, MPG, Filemaker, Visio, IThink, Access, MSProject, Screen, Paint, CHP, BMP, XMP. All these documents are stored in Fish Bins – see below | 335 of the digital documents are retained in their original hardcopy form in two archive boxes in Paul’s study. |
| 1. Fish Bins | Documents are stored by FISH in ‘bins’. Bins are created by the user and can have an unlimited number of documents within them. There are currently 91 bins named MO1, MO2, MO3, MO4, MO5, MO6, OL7, OL8, OL9 and so on to OL91 (the prefix MO was used initially when storage was on Magneto-Optical disks; when laptop storage capacity increased, files were stored on-line which is what the OL prefix stands for. As storage capacity on laptops continued to increase, the MO bins were themselves brought on-line but retained their original MO prefixes). | Bins are created in Fish by first creating a new directory in the desired location, and then navigating to that directory in the Fish Bin Setup screens. Fish places a control file called BIN.DAT in the directory. All Bins relating to the PAWDOC collection are stored in a directory in the My Documents section of TCL. Most bins contain documents amounting to between 200 and 600 Mb (200Mb was the most that an MO Disk could store – and it has been the aim to keep bins to around that size ever since). Total size of all Bins is currently about 40Gb. | None |
| 1. BT Cloud Backup software | This is a Cloud backup service which stores copies of specified documents or specified directories, to a cloud server. | The Cloud service and its associated software is provided by BT. The software resides on TCL in the BT Cloud directory with the path c:\Program Files | None |
| 1. Current bin Cloud Backup | The latest FISH bin (OL91) is currently being backed-up in the cloud, as have been OL76-81 and OL90. It was not intended that this service should store all bins – only the latest one. However, as older bins have been superceded by new ones, and as more storage in the service was provided, some of the older ones have been left in the cloud server. | These bins are stored in the BT Cloud Backup Service. | None |
| 1. Local Laptop Backup | This is a complete backup of all of the collection’s indexes, digital documents, and FISH databases. | Acer Aspire 5551 laptop. | The Acer Aspire 5551 laptop is located in a different room in the house |
| 1. Critical Files backup | This is a copy of the Index, the latest document Bin directory, the Fish file name control files, the Fish configuration management file, and the Fish databases. The copy is taken approximately once every few months. | The copy is made on a CD using the built in CD/DVD recorder in TCL. | The CD or DVD disk is stored in a drawer in a desk pedestal in Paul’s study. |
| 1. Bin backups on disk | A new disk is created for each group of 3 or 4 new bins. | The copies are made on DVDs using the built in CD/DVD recorder in the TCL. | The CD and DVD disks are stored in Paul’s wardrobe. |
| 1. Remote disc backup | This is a complete backup of the collection’s Indexes, digital documents, and Fish Databases. It is refreshed every few years. | The copies were made on DVDs using the built in CD/DVD recorder in the TCL. However they are now made on a portable hard drive. | The CD and DVD disks and portable hard drive are stored in the bedside table of the spare bedroom in Kirk Smeaton |
| 1. New Zealand backup | This is a complete backup of all of the collection’s indexes, digital documents, and FISH databases, as at March 2017. | The data is stored on a memory stick | The memory stick is stored in Matt’s house in Waitakere in NZ. |

1. **What are the current hardware and software platforms upon which the digital components operate? Are there any strategies or plans for the future evolution of these platforms?**

The hardware platform is an Acer Aspire 4830T Timeline X laptop (referred to throughout this document as The Collection Laptop – TCL) purchased at the end of 2011. It has an Intel Core i5 processor, 8Gb of RAM and a 750Gb hard disc. The operating system is Microsoft Windows 7. Attempts to upgrade to Windows 10 have failed and have now ceased.

The laptop will be replaced in the next few years with a more advanced model, probably running Windows 10 or its successor. However, a long term home is currently being sought for the collection, and, if one is found, it is not currently known what hardware and software platforms the destination repository will possess (see section 4 for further information about these requirements).

Despite the intentions stated above, it is also possible that any changes in the collection’s index and document management software platforms initiated by this Digital Preservation Project, may also present their own requirements for hardware and software infrastructure. Therefore, the hardware and software infrastructure strategy cannot be devised until the strategy for the collection’s indexing and document management software has itself been established.

1. **What risks do the different parts of the collection face? What actions should be taken to mitigate the risks? Who is responsible for each action?**
2. There is a risk that the Fish document management system will stop working and that the cost of getting it fixed and/or re-installed cannot be afforded.

Actions:

i) Document the FISH supplier’s recommended replacement route. Responsibility: Paul

ii) Document possible alternative Document Management systems / Database Systems and their costs. Responsibility: Jan

iii) Document any alternative solutions to using a Document Management System for storing and retrieving the collection’s electronic documents. Responsibility: Jan

1. There is a risk that the Fish document management system will be phased out and no longer supported and that the cost of purchasing and installing the recommended replacement cannot be afforded.  
   Actions: as for A. above
2. There is a risk that the current free version of MS SQL will not work on the next upgrade to the hardware and software platform  
   Actions: as for A. Above
3. There is a risk that it will be necessary to employ the Fish supplier, m-hance, to move Fish to the next planned hardware and software platform and that the cost of this service cannot be afforded.  
   Actions:  
   i) Identify actions once the options for replacing Fish have been investigated (see A. above). Responsibility: Paul
4. There is a risk that the current and future hardware and software platforms are incapable of opening some of the electronic documents contained in the collection.  
   Actions:  
   i) Run Droid across the collection’s documents and create a spreadsheet documenting the problem areas highlighted by the Droid results. Responsibility: Paul with assistance from Ross  
   ii) Decide what should be done to address the problems identified in the Droid analysis. Responsibility: Paul with assistance from Ross  
   iii) List the documents that don’t currently open and categorise them. Responsibility: Paul   
   iv) Decide what should be done for each category of document that doesn’t currently open. Responsibility: Paul with assistance from Ross/Jan  
   v) Decide what action, if any, should be taken on all categories of documents in the collection to promote their long term accessibility and survivability: Responsibility: Paul with assistance from Ross/Jan/Matt
5. There is a risk that the CD and DVD disks that are part of the collection, and the CDs and DVDs on which some of the backups for the collection are contained, may eventually become unreadable.  
   i) List the CDs and DVDs concerned and categorise them. Responsibility: Paul   
   ii) Decide what should be done for each category. Responsibility: Paul with assistance from Ross/Jan/Matt
6. There is a risk that the electronic element of the collection could become separated and lost from the hardcopy element of the collection.  
   Actions:  
   i) Document possible solutions and recommend a course of action. Responsibility: Paul
7. **List all the activities that you will need to do BEFORE you are in a position to create a realistic plan for the digital preservation work that is needed.**

* Decide what document management system or alternative, and any associated databases, are to be used going forward.
* Decide if Filemaker is to be retained as the platform for the Index or if it is to be replaced going forward.
* Establish the future platform strategy.
* Research and understand the actions required to make any moves planned from one piece of software to another; or from one platform to another.
* Research and understand the actions that are to be taken to be able to open those documents that don’t currently open.
* Research and understand the actions that are to be taken to promote the long term accessibility and survivability of all categories of document in the collection.
* Research and understand the actions that will need to be taken to mitigate against the collection’s CDs and DVDs becoming unreadable.
* Research and understand what actions will need to be taken to mitigate against the electronic part of the collection being separated from the physical part.

1. **What planning documents do you intend to produce after all the pre-plan activities defined in the previous answer have been completed, in order to manage the digital preservation work.**

* Digital Preservation Project Plan Description
  + Introduction
  + Background and Objectives
  + Scope (CLOPDAT)
  + Principles, Assumptions, Constraints and Risks
  + Governance
  + Milestones and Deliverables
  + Plan
  + Budget and Costs
* Digital Preservation Project Plan Chart

1. **What are your constraints in terms of cost and resources?**

* There is currently no money set aside for any software purchases or supplier support services.
* Specialist expertise on the project is voluntary and, in some cases, is having to be provided alongside full time jobs
* The team has little specialist knowledge or expertise of document management systems in general nor of Fish in particular

1. **How will the plans be validated? Who will be responsible for validating them?**

The team as a whole will validate the plans. Particular elements of the plan may also be passed to knowledgeable third parties for comment.

1. **How and when will the preservation arrangements be reviewed and updated, and who will do that work?**

The Preservation Project Plan will include an activity to produce a Preservation Maintenance Plan. The Maintenance Plan will define when the preservation arrangements are to be reviewed in the future and what actions should be taken at that time to update them as necessary. It will be the owner’s responsibility to implement the Preservation Maintenance Plan.   
**ENDS**

**Notes on changes to the template in the course of preparing this document (a revised process and associated templates will be documented in a paper at the end of the project)**

1. Change the title of 3) from ‘Why do you want to keep this data?’ To ‘Why do you want to keep this collection?’.

2. Move sections 5 (Why do you want to keep this data?) and 6 (For whom are you keeping it? etc.) to after Section 2 – this is a more logical flow of the contents.

3. Change the title of section 6 from ‘For whom are you keeping it? How are you going to test their expectations?’ to ‘For whom are you keeping it? What are their functionality, technology and any other requirements? How are you going to test their expectations?**’**

4.Combine 7 and 8 so that actions are directly under the risks. The way it is at present it is necessary to keep referring back to section 7 to see what the risk is you are talking about in 8**.**

5.Remove point 12 (What are your expectations of quality) – it is too broad a question at this early stage of the project. The achievement of quality standards for specific activities needs to be built into the project plan itself.