**PAW/DOC Content Check (Sep2015 – May2016)**

**By Paul Wilson**

**31May2016**

**Contents**

1. Introduction
2. Notation
3. Statistics
4. Problem areas
5. CD conversion
6. Digital inventory
7. Physical inventory

Appendix A - Detailed search results used to derive summary statistics

References

1. **Introduction**

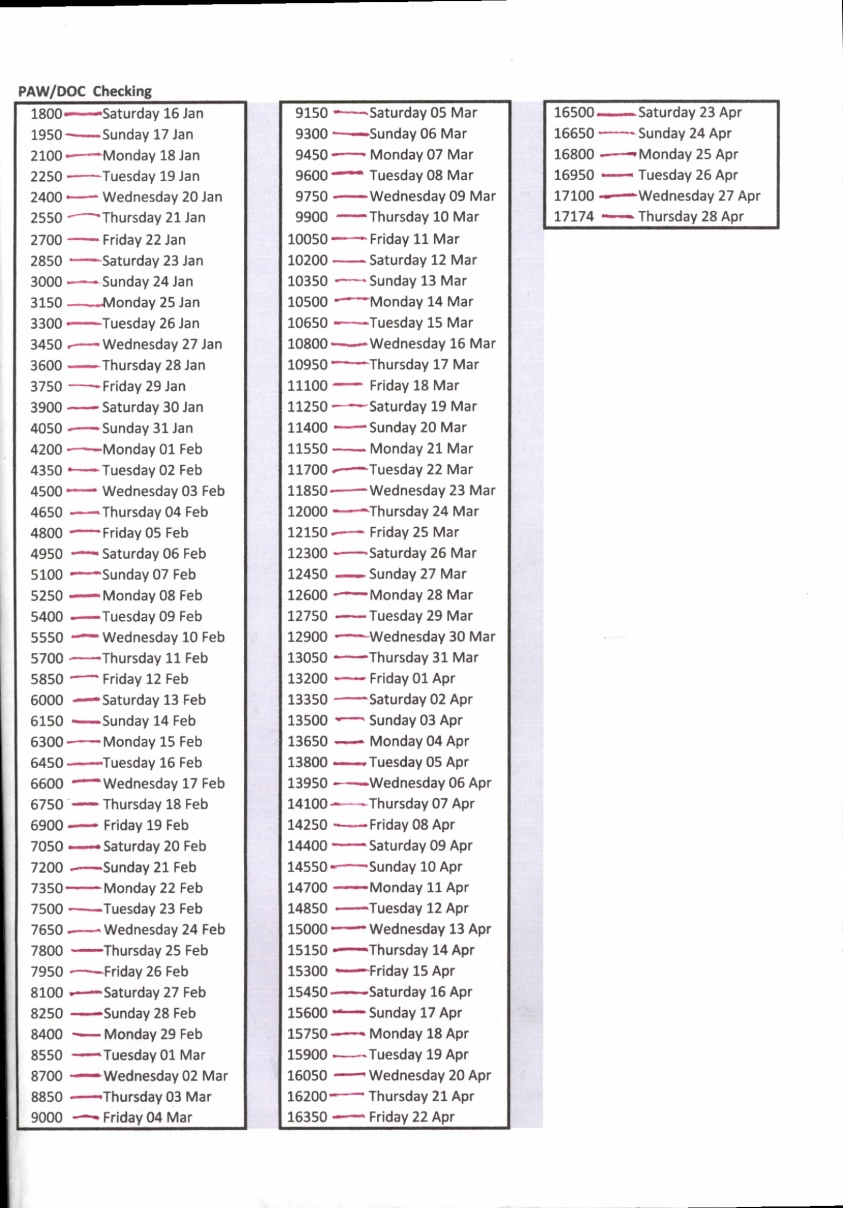
The name PAW/DOC refers to Paul Wilson’s filing system collection as documented in [Wilson, 2001]. It comprises an Index in a Filemaker database and a set of digital files held in the FISH Document Management System.

In 2015, the owner was preparing to undertake digital preservation work on PAW/DOC, and was also starting to make serious attempts to find a permanent repository for the collection [Wilson, April 2016]. It was anticipated that both these activities would require some kind of inventory check and tidying up of the collection. In September 2015, work started on digitising the small remaining subset of documents that were still only in physical form; and it was decided to embark on the checking activity as part of this final digitisation work.

In order to label items that had already been checked, it was decided to put either OK or XX in the Movements field of each Index entry. OK indicates that all is well with the Index entry and its associated collection of documents in the FISH Document Management System. XX indicates that there is a problem of some sort.

As the checking proceeded, a number of characteristics of OK entries were identified as well as various different types of XX problems. A notation was developed to record these characteristics/problems as qualifiers in brackets after OK or XX respectively. For example, OK(n docs) indicates that there is more than one document for this entry in the associated folder in Fish where n is the number of documents; and XX(lost) indicates that the physical document has been lost and there is no electronic version either for this entry.

In the course of the final digitisation work, about 400 index entries were checked and much of the notation was established. When the digitisation work was completed early in January 2016, the checking exercise started in its own right from the beginning of the Filemaker Index. It was time-consuming and laborious and was taking between one and two hours for every hundred index entries. I decided to set myself a target of 150 entries a day and duly created the record sheet in Figure 1 to prompt myself to keep to schedule (the inclusion of the record sheet in this document serves no purpose other than to provide a tangible token from which I can draw satisfaction from completing a very onerous task). It worked – I stuck to the schedule come what may (mostly working from about 6 am each morning) and finished some 3 months later. Some additional notations were identified over this period.



**Figure 1 – Checking Record Sheet**

I then set about checking the OK and XX labelling I had allocated, ensuring that every record had a label, analysing the number of instances of each characteristic and rectifying any problems that I could. I also zipped the contents of as many of the 40 or so CDs in the collection, included them in FISH, and destroyed the ones that were no longer needed. For those physical CDs and DVDs that I wanted to keep in the collection, I devised a pocket system that would store up to 8 discs in the fold down flap front of each of the two archive filing boxes that the remaining physical paper documents are stored in – see Figure 2 below. Having completed this exercise, all physical items in the collection were contained in the two archive boxes.



**Figure 2 – Discs stored in the flap front of an archive box**

1. **Notation**

Every entry in the Index has either OK or XX at the beginning of the Movements field to indicate that the Index entry and its associated FISH folder have been checked. Characteristics and Problems are recorded in brackets after OK or XX according to the Notation listed below. Note that any of the OK Characteristics may apply to both OK and XX entries; but that XX Problems only apply to XX entries.

| **Notation** | **Meaning** |
| --- | --- |
| ***OK Characteristics*** |  |
| app-x yyy | The FISH folder associated with this Index entry includes an Application file other than tif, pdf or jpg, where x is the number of instances of the application and yyy is the name of the application. Additional applications are appended after a semi-colon separator, eg app-3 Word; 1 PowerPoint. |
| cdde | The CD or DVD was destroyed at the request of the employer when I left, so there are no physical CD or electronic files left in the collection - only the index entry. |
| cddf | The CD or DVD was destroyed after its contents were copied into FISH. |
| cdro | The CD or DVD was removed by the Owner |
| dbl | One or more of the scanned files in FISH has double-sided pages in the wrong order (i.e. one side of a pile of documents was scanned and then the pile turned over and all the other sides were scanned). These are usually identified in the relevant FISH entries by the term "dbl unsorted". |
| empty | This index entry was inadvertently left empty when it was created and now serves no purpose other than confirming that the valid reference number is not missing. |
| left | The hardcopy document was left at a previous employer when I stopped working there, so there is no physical or electronic copy of this document - just the index entry. |
| multi | Contains one or more FISH electronic files in the form of multiple tif files - one file for each page. This is how FISH stores documents that are scanned under its control. |
| n docs | More than one document is held in Fish where n (which can be 2 or greater) is the number of different documents. If only 1 document is held in FISH there is no ‘n docs’ entry |
| ord | The pages of one or more of the scanned files in FISH are in the wrong order for a reason other than dbl (double-sided pages being scanned first one side and then the other). |
| phys | The physical original of this document, floppy disc, CD or DVD is in one of the Collection’s two archive boxes. Multiple instances of 'phys' indicate there are multiple separate physical items. |
| poor | The quality of one or more of the scanned electronic files in FISH is poor. |
| pro | The paper was removed from the collection by the Owner for his own personal use so there is probably no physical or electronic version of this item in the collection |
| refn | There is something wrong with the record number - it is duplicated or in some other way incorrect - but it does not affect the validity of the index entry. |
| sold | The CD or DVD was sold so is no longer in the collection |
| Str | One or more of the scanned electronic files in FISH have parts of the page showing as an image stretched down the page (for example, a line of text may appear to be stretched downwards for a few centimetres). This occurred when my first and second scanners were unable to feed the pages through properly, and it happened mostly with A5 documents, though there is at least one instance of it happening to an A4 document. |
| ***XX Problems*** |  |
| apperr-yyy xn | A file in FISH won't open correctly in the appropriate application. where yyy is the name of the application and n is the number of instances (eg. apperr-PowerPoint x3). |
| cont | Some of the content in an electronic file in FISH is missing (i.e. pages are missing) or is seriously flawed in some way. |
| disc | This item indicates that there is a floppy disc, CD or DVD which has yet to be copied to another medium to preserve its contents, for any of a variety of reasons including protection on the disc, disc won’t load, or not having the correct disc drive hardware. |
| emiss-yyy xn | A FISH electronic file is missing, where yyy is the name of the application and n is the number of instances (eg. emiss-Word x2). |
| ind | The Filemaker Index entry is blank or duplicated or there is some other serious problem with it. |
| lost | The hardcopy document has been lost and there is no electronic version in FISH either. |

1. **Statistics**

The number of occurrences of OK, XX, and associated Characteristics and Problems, in the Collection’s Index (as at 30May2016 when it contained 17206 entries) is shown in the table below (note that the numbers below probably contain some checking and recording errors; note also that the contents of some of the larger ZIP files are not included in these numbers; finally, note that the collection is still being added to so these numbers will increase going forward).

| **Notation** | **Search term to use in Filemaker Movements field** | **No of entries this appears in** | **Total No of instances within all entries** | **Notes** |
| --- | --- | --- | --- | --- |
| **OK** | OK | 16751 | n/a |  |
| app-xxx | app- | 3494 | 8123 | Total instances calculated from instances of each app |
| app-n Word | None | 3103 | 6380 | Calculated from detailed search results - see App A |
| app-n PowerPoint | None | 347 | 466 | Calculated from detailed search results - see App A |
| app-n Excel | None | 215 | 625 | Calculated from detailed search results - see App A |
| app-n HTML | None | 119 | 382 | Calculated from detailed search results - see App A |
| app-n Help | None | 4 | 90 | Calculated from detailed search results - see App A |
| app-n zip | None | 46 | 92 | Calculated from detailed search results - see App A |
| app-n mpg | None | 20 | 20 | Calculated from detailed search results - see App A |
| app-n fmpro7 | \*"fmpro" | 14 | 36 |  |
| app-n visio | \*"visio" | 8 | 13 |  |
| app-n ithink | \*"ithink" | 2 | 2 |  |
| app-n access | \*"access" | 2 | 5 |  |
| app-n msproject | \*"msproject" | 1 | 1 |  |
| app-n screen | \*"screen" | 1 | 4 | This is Lotus ScreenCam |
| app-n paint | \*"paint" | 2 | 3 |  |
| app-n chp | \*"chp" | 1 | 1 |  |
| App-n bmp | \*"bmp" | 1 | 2 |  |
| app-n xmp | \*"xmp" | 1 | 1 |  |
| cdde | \*"cdde" | 34 | 34 |  |
| cddf | \*"cddf" | 37 | 37 |  |
| cdro | \*"cdro" | 25 | 25 |  |
| dbl | \*"dbl" | 622 | Unknown |  |
| empty | \*"empty" | 5 | n/a |  |
| left | \*"left" | 16 | 16 | 4 Ref items identified in the search were excluded |
| multi | \*"multi" | 8606 | Unknown |  |
| n docs | None | 17205 | 28418 | Calculated from detailed search results - see App A |
| ord | \*"ord" then omit \*"word" | 4 | Unknown |  |
| phys | \*"phys" | 335 | Unknown |  |
| poor | \*"poor" | 53 | Unknown |  |
| pro | \*"pro" then omit \*"fmpro" | 91 | n/a |  |
| refn | \*"ref" | 4 | n/a |  |
| sold | \*"sold" | 3 | n/a |  |
| Str | \*"str" | 25 | Unknown | It is likely that several other instances of this characteristic were missed during the check. |
| **XX** | \*"XX(" | 455 | n/a | This is not the total of all items below because there can be multiple ‘apperr’ in one entry. |
| apperr-xxx | \*"apperr" | 162 | 296 | The number of instances total was calculated from all the entries below |
| Apperr-Word | \*"apperr-word" | 12 | 22 |  |
| Apperr-PowerPoint | \*"apperr-PowerPoint" | 126 | 155 |  |
| Apperr-Excel | \*"apperr-excel" | 1 | 1 |  |
| Apperr-Visio | \*"apperr-visio" | 4 | 6 |  |
| Apperr-Access | \*"apperr-Access" | 2 | 5 |  |
| Apperr-Help | \*"apperr-Help" | 2 | 87 |  |
| Apperr-pdf | \*"apperr-pdf" | 2 | 2 |  |
| Apperr-tif | \*"apperr-tif" | 1 | 1 |  |
| Apperr-bmp | \*"apperr-bmp" | 1 | 2 |  |
| Apperr-paint | \*"apperr-paint" | 1 | 1 |  |
| Apperr-xml | \*"apperr-xml" | 1 | 1 |  |
| Apperr-html | \*"apperr-html" | 3 | 4 |  |
| Apperr-msproject | \*"apperr-msproject" | 1 | 1 |  |
| Apperr-screen | \*"apperr-screen" | 1 | 4 |  |
| Apperr-ithink | \*"apperr-ithink" | 1 | 1 |  |
| Apperr-exe | \*"apperr-exe" | 3 | 3 |  |
| cont | \*"cont" | 6 | 6 |  |
| disc | \*"disc" | 6 | 7 |  |
| emiss-xxx | \*"emiss" | 3 | 22 | The number of instances total was calculated from all the entries below |
| emiss-tif | \*"emiss-tif" | 2 | 18 |  |
| emiss-excel | \*"emiss-excel" | 1 | 4 |  |
| ind | \*"emiss-ind" | 0 | n/a | Issues with these 10 items were resolved after the initial check had been completed |
| lost | \*"emiss-lost" then omit \*"OK" | 278 | Unknown |  |

1. **Problem Areas**

Problem areas are identified by ‘XX’ in the Movements field in the Filemaker Index. Statistics for these areas are provided above. This section describes the investigations that were done to understand the problems and why they occurred, and to try to rectify them if possible.

* 1. **Initial fixes**

Immediately following the checking, a number of the more obvious problems were fixed. Those are listed here to provide a checklist of some of the types of problems that can occur:

* **Text in FISH folder names:** 11 spurious folders appeared at the top of the folder list in FISH with text in their titles instead of a Reference Number. These had been produced erroneously because I had copied the wrong field in Filemaker when creating new folders in FISH. About 8 missing files were found in these spurious folders and placed into their correct folders.
* **Blank in FISH Folder name:** One item simply had an incorrect blank at the beginning of the FISH folder name which meant it wasn’t being identified when searching FISH for the Filemaker Index Reference Number. It was identified by using the more detailed FISH search function, and rectified.
* **File stored in adjacent FISH folder:** One item did actually have a file in FISH but it had been included in the folder of the following Ref No. It was moved into the correct folder.
* **Incorrect Filemaker Ref No:** One item had text in the Filemaker Reference number. It had been created in error and then left and another correct entry created. The Reference Number was given the missing number in the Reference Number sequence and linked to the relevant entry.  
    
  1. **APPERR items**

These are items for which a file exists in FISH but the file cannot be opened - for whatever reason - by the application set currently on the laptop. For each problem application one instance of the APPERR notation appears in the Filemaker Movements field together with an indication of the number of problem files that exists, for example, ‘apperr-PowerPoint x3’ indicates that there are three PowerPoint files in the relevant FISH folder which can’t be opened.

There are 296 problem apperr files in the collection of which 155 are PowerPoint files created in an early version of PowerPoint. When an attempt is made to open the file with a later version of PowerPoint a message is displayed saying *“PowerPoint can’t open the type of file represented by [filename]”*. Research on the net indicates that this is a well known problem with files created in PowerPoint 97 or earlier. Some people advise that the ZAMZAR site can perform a conversion, but it was not able to convert the test file that I sent it. Further research is required to identify a solution.

87 of the 296 apperr files are Help files, though this is misleading - 86 of the files are part of one single large help file. When trying to open these files, the following message is displayed *“The Help for this program was created in Windows Help format, which depends on a feature that isn't included in this version of Windows. However, you can download a program that will allow you to view Help created in the Windows Help format. For more information, go to the Microsoft Help and Support website.”*. Hovering over the link produces a message saying *“Go to support.Microsoft.com and search for WinHlp32.exe”*. Further work is required to obtain this exe and to try it out.

17 of the 296 apperr files won’t open simply because I haven’t got the relevant applications installed on my laptop (Visio, Access, MS Project, Lotus ScreenCam, and iThink).

No investigation has been performed into why the remaining 36 apperr files won’t open. They are of a variety of application types – 22 Word, 1 Excel, 2 PDF, 1 TIF, 2 BMP, 1 Paint, 1 XML, 4 HTML, and 3 Exe (all three of the exe problems are CD compilations that won’t open).

* 1. **Cont items**

These 6 items have something wrong with their contents. PAW/DOC/4012/08 has two pages missing from both the hardcopy (which has been retained) and the scanned versions - it looks as though they were originally mislaid in the paper version. It is possible that these could be retrieved from the City University library where this document (Charlie Chan’s dissertation on putting PAW/DOC onto optical disc) will probably be stored.

PAW/DOC/5269/01 and PAW/DOC/5500/02 are both two page scanned documents - and in both cases the two pages are completely black. It seems unlikely that other copies of these documents could be obtained – on is a brochure on Telford’s Postal services in the early 1990s; and the other is a flier on business process engineering books from the same era.

For one item (PAW/DOC/5524/02), all the content is present, but it contains an explicit instruction not to reproduce it – I suppose the owners should be asked if they will waive this requirement for this particular collection.

The fifth item (PAW/SLI/0027/01) has the wrong contents because the wrong overhead slide was scanned – it is unlikely that this can be retrieved unless another copy exists elsewhere in the collection.

The sixth item is a CD (PAW/CD/0004/01) from which it was only possible to copy a subset of its files into FISH - not sure why. Anyway the physical CD has been retained and is still working, so there is still time to try and understand what the problem is and to rectify it, before the data are lost.

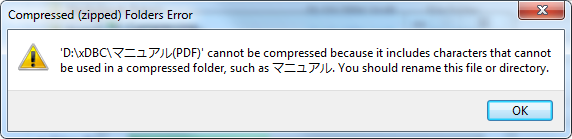
* 1. **Disc items**

Six floppy discs, CDs or DVDs have yet to be copied to another medium for a variety of reasons. Two are 3.5 inch floppy discs which require a floppy disc drive which I do not have (PAW/DOC/3637/01 and PAW/CD/005/01).

One is a DVD (PAW/DVD/005/01) which appears to possess some kind of protection mechanism which prevents it being copied (this must have been installed by the company that I employed to copy the video from VCR onto disc – probably because it was an IBM video and they didn’t want to risk having any kind of problem with IBM).

One is a FISH installation CD dated 04Feb2000 (PAW/CD/005/01) which just won’t load – it just keeps on spinning.

Two are Filemaker installation CDs which I was unable to zip (PAW/CD/026/01 and PAW/CD/032/01). When I tried to do so the message in Figure 3 below was displayed. There are too many files to store separately in FISH.



**Figure 3 – Message displayed when trying to Zip the contents of two of the CDs**

* 1. **Emiss items**

These are items for which an entry exists in a FISH folder but FISH cannot find the file associated with it. The checking identified about 150 such instances, however the majority of these items were found on the original 18 year old backup discs for the FISH bins MO4 and MO5 – an illustration of just how important backups are (note that the file names of MO4 and MO5 indicate this discrepancy – *‘MO4 - incomplete – should have 3856 files’* and *‘MO5 – incomplete – should have 3718 files’*; however now that the missing files have been recovered these words are no longer valid but they remain in the file names because FISH relies on the exact file title).

Having recovered the missing files there are now only 3 entries for which electronic files cannot be found in FISH:

PAW/DOC/4245/24 has the last three scanned pages of a 26 page document missing. This almost certainly resulted from corruption in the optical disc I was using in 1998 and so the pages are unlikely to be found.

PAW/DOC/4457/01 has pages 1-11 and 23-26 missing from the scan of a 46 page document. It is not known how this occurred but it is possible that it could have been another manifestation of the optical disc corruptions that occurred in 1998 and documented in PAW/DOC/6588-01 and 02.

Finally, four Excel documents are missing from PAW/DOC/1146/20. The FISH DoFilesExist utility, which checks for missing files and their presence in incorrect bins, provides an unusual report for these four items. It seems to indicate that FISH-specific file names were not provided for them unlike the other missing files which have names such as 63BC.tif. The DoFilesExist log report specifies the location and names of the four files in their original location on my laptop (eg. C:\APAW\FORMFILD\EXPENSES\%X11AP97.XLS). I don’t understand why this has occurred. It will be necessary to discuss this with FISH support personnel to try to understand what has happened.

* 1. **Ind items**

Ind items have a Filemaker Index entry which is problematic in some way, and the checking identified 10 such instances. Each one was investigated and remedial action taken as appropriate:

Five of the entries seem to have been created in error, but given valid Reference Numbers, and left empty – perhaps because I’d been distracted after the entry was created and had forgotten I’d created the entry when I got back to it. Since the Reference Numbers are valid, the classifications were changed from XX to OK(empty).

One of the entries is a duplicate of another entry with exactly the same Reference Number. Since the other entry has valid contents with associated electronic files in FISH, this entry was not needed so it was deleted from the Filemaker index.

One of the index entries contained a cryptic message in the Title field - '(original index entry deleted by mistake - should find it on earlier versions of PAWDOC)'. However it couldn’t be found in any of the earlier backups of the index, and the entry has a valid Reference Number and contents. Therefore its classification was changed from XX to OK.

One item had a completely incorrect Reference Number (PAW/BTJ/) but a valid set of contents and an associated FISH electronic file. So, the Reference Number was changed to the nearest available one in the sequence (PAW/DOC/8146/02), a new equivalent folder was created in FISH for the electronic file, and the classification was changed from XX to OK.

The final two items had spurious Reference Numbers (‘PAW/DOC/4’ and ‘P’), had no contents and there are no missing Reference Numbers in the sequence around them, so they were both deleted from the Filemaker Index.

As a result of these actions taken there are now no ‘ind’ items In the Index.

* 1. **Lost items**

The hardcopy of these items has been lost and there are are no electronic versions in FISH. The 278 items of this type fall broadly into the following five categories: 87 were related to mislaid journals, 34 were not under my control (they were from the collection of a colleague who worked for me, which I included in my own collection when he left); 16 were borrowed by other people and not returned; 9 were the result of not copying text correctly from emails, and there is no clear reason why the remaining 131 were lost. A detailed analysis is provided below:

1. **RMC/DOC items:** These are 34 files in the Richard McCullum document set which is included as a whole in the collection (Richard worked for me briefly around 1988). I don't know why these were missing - but it could be that they only existed as electronic documents and I think I only acquired the hardcopy documents – and no electronic documents - from Richard.
2. **PAW/HSM items:** Of the remaining 244 items, 5 are Human Systems Management Journal articles. I think I must have just thrown these journals away at some point without realising I had indexed some articles within them. I did manage to find some of the missing HSM items on the net, but payment was required to obtain them, so I just placed the URL for their abstracts in the Movements field and left them classified as XX(lost) items.
3. **PAW/BTJ items:** Of the remaining 239 items, 6 are British Technology Journal articles. Like the HSM journal, I think I must have thrown these away at some point without realising I had indexed some articles within them. I tried to find them on the net with no success, so they remain classified as XX(lost).
4. **PAW/CIN item:** Of the remaining 233 items, 1 was an article from the Creativity & Innovation Network journal. There are 42 other CIN entries in the index, all with the equivalent text in FISH, so I must have made a mistake when scanning the hardcopy journals and missed this one article. I made a quick search on the net for it but couldn't find it, so I it remains classified as XX(lost).
5. **PAW/Fiche item:** Of the remaining 232 items, 1 was a microfiche. I don't know where this went - and the entry in the PAW/SET index just says *'not known'* for Physical Set location, so I have left this item as XX(lost). However, in looking for the fiche I checked out my old microfiche storage booklet in the loft and found several other fiche from my days in Distribution Planning at Kodak and two others; so I added these into the collection as PAW/Fiche02-05, mounted them on A4 pages, and placed them in the Physical artefact archive boxes. Note that there is no electronic version or backup of what is on these microfiche.
6. **MACUK items:** Of the remaining 231 items, 15 were MacUser UK magazine articles (fourteen of them with the Ref No PAW/MACUK, and one with the Ref No PAW/DOC/MACUK). I think I must have just thrown these away at some point without realising I had indexed some articles within them. I tried to find them on the net with no success, so they remain classified as XX(lost).
7. **MACTIM items:** Of the remaining 216 items, 9 were from three issues of the UK Macintosh User Group Mac Times magazine (seven of them with the Ref No PAW/DOC/MACTIM, and two with the Ref No PAW/MACTIM – I clearly had problems retaining consistency in Ref No naming). Several other items from other issues of this magazine have been successfully scanned and included in FISH, so I think I must have mislaid these issues sometime before I started scanning. I had no luck tracking them down on the net so they remain classified as XX(lost).
8. **MACUSA items:** of the remaining 207 items 1 was from the Feb90 issue of the USA Macintosh magazine (with Ref No PAW/DOC/MACUSA). There are over 20 other articles from other issues of MACUSA in the collection, so I think I must have mislaid this issue sometime before I started scanning. I had no luck tracking it down on the net so the item remains classified as XX(lost).
9. **PAW/CSCW items:** Of the remaining 206 items, 25 were from the CSCW journal. Five of these were from just two issues of the journal, and the Movements fields record that they were borrowed by groups of individuals; the assumption must be that they weren't returned. I suspect that is also what happened to the other missing items - even though they have no 'B' (for Borrowed) signifier in the Movements field. The articles are available on the net but all require payment, so these items remain classified as XX(lost).
10. **PAW/INTER item:** of the remaining 181 items, 1 was an article from the ACM Interactions magazine from April 1998. I had successfully downloaded 4 other items from that issue when I was a member of ACM in 2007, but must have just missed downloading this item in error. Now that I'm not a member, I was able to find the article on the ACM website but they wanted to charge me $15 to download it. So I just put the URL in the movements field and left the item classified as XX(lost).
11. **PAW/ACMCOM items:** Of the remaining 180 items, 1 was an article from the Communications of the ACM Journal. As noted in the movement field, I had mistakenly imported the file for the previous entry (on the subject of reciprocity) when I was uploading from the ACM library and obviously didn't check. I was able to find the article on the ACM website but, now that I'm not an ACM member, they wanted to charge me to download it. So I just put the URL in the movements field and left the item classified as XX(lost).
12. **PAW/JOC items:** of the remaining 179 items, 21 were papers from the Journal of Organisational Computing. Five of these were from a single issue of the journal (Ref PAW/JOC/Sep96/...) and the Movements field records that it was borrowed by a group of individuals; the assumption must be that it wasn't returned. I suspect that is also what happened to the other missing JOC items - even though they have no 'B' signifier in the Movements field. Consequently I have left these items classified as XX(lost), but have included the URLs of the relevant issues of the journal in the Movements fields of all of the missing items.
13. **PAW/TM items**: of the remaining 158 items, 7 were papers from an unknown journal. The only clue as to what it was is the Ref No (PAW/TM/2.2A/...). However, the PAW/Set index was obviously constructed after this item had disappeared - it says *"Don't know what this is - but it looks like a journal from the entries in PAW/DOC (could it be "Telecommunications Management"?)"*. Despite my best efforts at searching the net, I have been unable to identify what publication they refer to. Consequently, these items remain classified as XX(lost).
14. **Text not copied correctly from email:** 9 of the remaining 151 items were Word documents which were supposed to contain the text of email messages but for which the text had not been copied correctly in the email system and which I had failed to check after I had created the document (this probably occurred because of the vagaries of the Lotus Notes email system). The Word documents in question are either completely empty or just have a few words at the beginning. I have left these items classified as XX(lost – text not copied correctly) – though it should be noted that in the case of PAW/DOC/1199/26 there are 25 other Word documents associated with that Ref No in FISH that don’t have this problem and that have their contents fully intact.
15. **Borrowed:** of the remaining 142 items, 11 had been borrowed by individuals, or circulated round groups, and not returned. These remain classified as XX(lost). Note that when doing a search for these items (using the search term **lost B(** in the Movements field), 16 items are found. However 5 of these are for journals (CSCW and JOC) which have already been taken into account earlier in this section under the relevant journal heading.
16. **EF or scanned items:** Of the remaining 131 items, 11 had abbreviations (such as EF- Electronic File, S – scanned, or PD – paper destroyed) in the Movements field indicating that an electronic file existed. On checking out the four with S and/or PD in the Movements field, one (PAW/DOC/4245/23) appears to have been lost in the optical disk corruption around 1998 as fully documented in PAW/DOC/6588/02. The other three (PAW/DOC/4011/75, PAW/DOC/6070/01 and PAW/DOC/7286/01) have S or PD in the Movements field which indicates that they were scanned, however there are no files in FISH and a search of the FISH description field failed to find the documents concerned. The other 7 items (PAW/DOC/6201/01, 7077/01, 7325/01, 7857/01, 7970/01, 8491/05 and 8876/01) all have EF in the Movements field, however there are no files in FISH and a search of the FISH description field failed to find the documents concerned. Consequently all these 11 files remain classified as XX(lost).
17. **Lost after archiving:** Of the remaining 120 items, 17 had the classification A(date) in the movements field indicating that they had been archived before being found to be lost. Assuming that they were not removed from the archive boxes until they were scanned (though I can’t be 100% sure of that), they were most likely lost in the middle of the bulk scanning process. These retain their classification of XX(lost).
18. **Lost before archiving:** Of the remaining 103 items, 35 had the classification L in the Movements field without any other classification such as A (for archiving) or B(for borrowed), indicating that it was the hardcopy had been mislaid during the normal course of filing (to search for these items, first search for **lost** in the Movements field, then refine the search by omitting **A(** and then by omitting **B(** ). These retain their classification of XX(lost).
19. **Qualifying explanations:** Of the remaining 68 items, two have qualifying explanations in the Movements field: One (for PAW/DOC/3442/01) is *'Destroyed in error during Jan94 sort out at Weybridge'* - can't remember how this happened; and the other (for PAW/DOC/9005/01) is *'Waiting for Fish to be installed on my machine'* - which sounds like it was probably just an electronic file which I forgot to load into FISH i.e. there should probably have been an EF in the Movements field. Both of these remain classified as XX(lost).
20. **Document in adjacent entry:** Of the remaining 66 items, one (PAW/DOC/7222/02) appears to have been created before 7222/01 which is a little bit odd. But, ignoring this anomaly, I think I found the document it referred to elsewhere in the 7222 set of documents in FISH (I don't have a record of exactly what I did). So I moved that document into the correct FISH folder and changed the classification of 7222/02 to OK.
21. **No explanation:** There is no obvious explanation for the loss of the final 65 items. I did a search in FISH for phrases from the titles of each of these items but didn't come up with any hits. It is possible that some of the items created after 2000 were electronic files and just didn't have EF placed in the Movements field. However, this doesn’t explain why they aren't in FISH. For the earlier items that were originally paper, it is probable that they were lost before I got to scan the particular box they should have been in. However, this is all conjecture. All of these items remain classified as XX(lost).
22. **CD conversion**

During the main checking exercise, PAW/CD index entries (but not PAW/DVD entries which already had FISH entries) were bypassed and instead were dealt with during the subsequent analysis exercise. At that point it was decided to not just allocate appropriate classifications, but also to a) include the CD’s electronic files into FISH where possible, b) to destroy unwanted physical CDs and DVDs, and c) to place the physical CDs and DVDs that were being retained into the document archive boxes.

1. **CD copying and zipping**

Details of how the contents of each CD were copied and placed into FISH are contained in section 18 of the Checking Working Document [Wilson, May 2016]. Most of the CDs contained large numbers of files which would be impractical to store individually in FISH, so a single ZIP file was created for each of these CDS which was then stored in FISH.

There are 43 PAW/CD Index entries and 36 were given an OK classification. 27 of these OK items have been removed from the collection mainly because the hardware or software they relate to is no longer being used (these items have the **CDRO** or **sold** notations in their Movements field). For these items there are no FISH entries. Of the 7 items given an XX classification, four CDs wouldn’t display their contents or couldn’t be copied correctly for some other reason (PAW/CD/004/01, 005/01, 026/01, 032/01) and will require preservation work doing on them at some point (these items all have **disc** in their Movements field). Two of the others (PAW/CD/002/01 and 004/01) have been zipped and stored in FISH satisfactorily, but, once the zipped files are extracted, the initiating executable won’t run to provide the index to the contents (these items have **apperr-EXE** in their Movements field). The final XX item (PAW/CD/008/01) contains an older PowerPoint file that won’t open (this has apperr-PowerPoint in the Movements field).

For the PAW/DVD collection, all the items have an OK classification other than PAW/DVD/005/01 which can’t be copied from the CD (this XX item has **disc** in the Movements field).

1. **Destruction of physical CDs and DVDs**

CDs and DVDs which had copied successfully into FISH and which had no special CD label, or were not associated with a software licence, were destroyed or moved to the backup collection. The remaining CDs were taken out of their cases and prepared for storage in the archive boxes.

1. **Storage of physical CDs and DVDs in archive boxes**

A plastic wallet system was constructed to fit into the door flaps of each archive box. Each flap has 8 wallets – so overall there is the potential to store 16 discs. However, with the archive boxes standing on the floor, the flap of the lower box won’t hang down vertically, so, in this configuration, the top three wallets in the lower flap are not used. At present, thirteen physical discs are being stored in the archive boxes – 11 of them from the PAW/CD collection and 2 of them from the PAW/DVD collection. In one instance (PAW/CD/005/01) a 3.5 inch floppy disc is also stored with the CD in the same wallet.

1. **Digital Inventory**

The PAW/DOC collection refers to all physical and electronic components related to **all** document sets listed in the PAW/SET index. Digital components refer to all electronic files associated with the collection. Physical components relate to all other components. Note that scanning equipment and software, and application programmes to read the stored files, are not included in these inventories. The electronic components in the PAW/DOC collection are listed in the table below:

| **Component** | **Path or Location** | **File name(s)** | **Notes** |
| --- | --- | --- | --- |
| E1. Set Index | C:\Users\Paul\Documents\My Documents\ A - Master Index Documents | Pawset.fp7 | This is a Filemaker database which records all the subsets of documents in the collection’s Index eg. PAW/DOC. PAW/CD, PAW/CHI, RMC/DOC, NCC/CJA. |
| E2. Document Index | C:\Users\Paul\Documents\My Documents\ A - Master Index Documents | Pawdoc.fp7 | This is a Filemaker database containing the main Index of the collection. It is a key file and needs regular backup. |
| E3. Fish software | C:\Maxima\FiShServer | Fish.exe + many others + subfolders | This is the document management software used to store the electronic files associated with the Index entries. |
| E4. Fish file name control | C:\Maxima\FiShServer\Storage | Highest.doc + many others | This needs regular backup its contents change as new files are stored in FISH. |
| E5. Fish Configuration File | C:\Maxima\FiShServer | PCLIPWIN.INI | A new version with a new date is saved each time FISH is used so it is backed up regularly. |
| E6. Fish Databases | C:\Maxima\FiShDatabaseFiles | fish\_Data.MDF + fish\_Log.LDF | These are key SQL databases and need regular backup. |
| E7. Fish Upgrade Info | C:\Maxima\Media | Various files in sub-folders | Files associated with the FISH upgrade on the laptop in Jan 2012. |
| E8. Fish documents | C:\Users\Paul\Documents\My Documents\Fish OL folders | Over 90 sub-folders | Each sub-folder is a defined FISH ‘bin’. Total file size is over 40Gb. |
| E9. BT Backup software | C:\Users\Paul\AppData\Local\F-Secure\BT Cloud | BT Cloud.exe and others | BT Cloud service will provide a new copy of the Cloud software if required. |
| E10. Current bin Backup | In BT Cloud | The name of the current Bin eg. OL90 | Need to remember to include each newly created bin in BT Cloud |
| E11. Local Laptop Backup | In the other laptop in the house in the following directory: C:\Users\Su\Documents\My Documents\ Paul’s Backups | Components 1-8 | This backup is refreshed every few years |
| E12. Critical Files backup | Older versions are on disc; current version is on the Bipra portable hard drive | Components 1,2,4,5,6,8 | This backup is refreshed every few months |
| E13. Bin backups on disc | Stored elsewhere in the house | Component 8 | A new disc is created for each group of 3 or 4 new bins |
| E14. Remote disc backup | Stored over 100 miles away | Components 1,2,4,5,6,8 | This backup is refreshed every few years |
| E15. New Zealand backup | Stored in New Zealand | Components 1,2,4,5,6,8 | These contain the data as at Jan 2012 |

1. **Physical Inventory**

The physical components associated with the collection are listed in the table below.

| **Components** | **Location** | **Description** | **Notes** |
| --- | --- | --- | --- |
| P1. Collection Computer | Desk in Paul’s study | Acer Aspire 4830T laptop with CD/DVD drive |  |
| P2. Local backup computer | Desk elsewhere in the house | Acer Aspire 5551 laptop |  |
| P3. Bipra hard disc | Pedestal in Paul’s Study | Bipra 500Gb portable hard disc drive |  |
| P4. Archive Boxes | Paul’s study | Two stackable, front opening, storage boxes |  |
| P5. Physical documents | In the Archive Boxes | The physical documents that were retained after being digitised |  |
| P6. Physical CDs and DVDs | In the flap fronts of the Archive Boxes | The physical CDs and DVDs that have been retained |  |
| P7. Critical Files backup discs | Disc box in the Pedestal in Paul’s Study; and Paul’s wardrobe 2nd shelf | These contain copies of the critical files taken on various dates from 2006 onwards |  |
| P8. Bin backups on disc | Paul’s wardrobe, 4th shelf | There are two sets of discs: one contains copies of all bins after they were filled and superceded by a subsequent bin; the other is a copy of bins 1-52 taken in 2008. |  |
| P9. Remote disc backup | In a drawer in Yorkshire | This is a set of discs in plastic sleeves |  |
| P10. New Zealand backup | In New Zealand | This is a set of discs in plastic sleeves |  |
| P11. Disaster Recovery file | Pedestal in Paul’s Study | This is a plastic folder containing paper documents recording backup arrangements and information that would assist in recovering the PAW/DOC system. |  |

**Appendix A - Detailed search results used to derive summary statistics**

The table below provides the detailed search results used to derive statistics associated with the results of the checking exercise.

| **Notation** | **Search term to use in Filemaker Movements field** | **No of entries this occurs in** | **Total No of instances within all entries** | **Notes** |
| --- | --- | --- | --- | --- |
| **OK** | OK | 16751 | n/a |  |
| app-n yyy | app- | 3494 | 8123 | Total instances calculated from instances of each app |
| app-Word | None | 3103 | 6380 | Calculated by adding the results of the searches on Word documents below |
| app-1 Word | \*"1 Word" | 2161 | 2161 |  |
| app-2 Word | \*"2 Word" | 444 | 888 |  |
| app-3 Word | \*"3 Word" | 184 | 552 |  |
| app-4 Word | \*"4 Word" | 95 | 380 |  |
| app-5 Word | \*"5 Word" | 54 | 270 |  |
| app-6 Word | \*"6 Word" | 36 | 216 |  |
| app-7 Word | \*"7 Word" | 32 | 224 |  |
| app-8 Word | \*"8 Word" | 14 | 112 |  |
| app-9 Word | \*"9 Word" | 17 | 153 |  |
| app-10 Word | \*"10 Word" | 13 | 130 |  |
| app-11 Word | \*"11 Word" | 3 | 33 |  |
| app-12 Word | \*"12 Word" | 5 | 60 |  |
| app-13 Word | \*"13 Word" | 6 | 78 |  |
| app-14 Word | \*"14 Word" | 5 | 70 |  |
| app-15 Word | \*"15 Word" | 3 | 45 |  |
| app-16 Word | \*"16 Word" | 3 | 48 |  |
| app-17 Word | \*"17 Word" | 1 | 17 |  |
| app-18 Word | \*"18 Word" | 1 | 18 |  |
| app-19 Word | \*"19 Word" | 1 | 19 |  |
| app-20 Word | \*"20 Word" | 3 | 60 |  |
| app-21 Word | \*"21 Word" | 3 | 66 |  |
| app-22 Word | \*"22 Word" | 0 | 0 |  |
| app-23 Word | \*"23 Word" | 3 | 69 |  |
| app-24 Word | \*"24 Word" | 1 | 24 |  |
| app-25 Word | \*"25 Word" | 1 | 25 |  |
| app-26 Word | \*"26 Word" | 1 | 26 |  |
| app-27 Word | \*"27 Word" | 0 | 0 |  |
| app-28 Word | \*"28 Word" | 1 | 28 |  |
| app-29 Word | \*"29 Word" | 0 | 0 |  |
| app-33 Word | 3# Word | 1 | 33 | Deduced from search results |
| app-43 Word | 4# Word | 1 | 43 | Deduced from search results |
| app-46 Word | 4# Word | 1 | 46 | Deduced from search results |
| app-48 Word | 4# Word | 1 | 48 | Deduced from search results |
| app-50 Word | 5# Word | 2 | 100 | Deduced from search results |
| app-51 Word | 5# Word | 2 | 102 | Deduced from search results |
| app-52 Word | 5# Word | 1 | 52 | Deduced from search results |
| App-53 Word | 5# Word | 2 | 106 | Deduced from search results |
| App-78 Word | 7# Word | 1 | 78 | Deduced from search results |
| **Total Word docs** |  | **3103** | **6380** | **Sum of all the above** |
| app-PowerPoint | None | 347 | 466 | Calculated by adding the results of the searches on PowerPoint documents below |
| app-1 PowerPoint | \*"1 PowerPoint" | 277 | 277 |  |
| app-2 PowerPoint | \*"2 PowerPoint" | 43 | 86 |  |
| app-3 PowerPoint | \*"3 PowerPoint" | 19 | 57 |  |
| app-4 PowerPoint | \*"4 PowerPoint" | 3 | 12 |  |
| app-5 PowerPoint | \*"5 PowerPoint" | 2 | 10 |  |
| app-6 PowerPoint | \*"6 PowerPoint" | 1 | 6 | One item has a 6 and a 3 = 9 |
| app-7 PowerPoint | \*"7 PowerPoint" | 0 | 0 |  |
| app-8 PowerPoint | \*"8 PowerPoint" | 0 | 0 |  |
| app-9 PowerPoint | \*"9 PowerPoint" | 2 | 18 | One item identified in \*"6 PowerPoint" search |
| app-10+ PowerPoint | 1# PowerPoint | 0 | 0 | Confirmed by checking each of the 24 hits |
| **Total PowerPoint docs** |  | **347** | **466** | **Sum of all the above** |
| app-Excel | None | 215 | 625 | Calculated by adding the results of the searches on Excel documents below |
| app-1 Excel | \*"1 Excel" | 147 | 147 |  |
|  | \*"2 Excel" | 32 | 64 |  |
|  | \*"3 Excel" | 16 | 48 |  |
|  | \*"4 Excel" | 6 | 24 | Identified a ’44 Excel’ instance |
|  | \*"5 Excel" | 2 | 10 |  |
|  | \*"6 Excel" | 5 | 30 | Identified a ’19 Excel’ instance |
|  | \*"7 Excel" | 0 | 0 | Identified a ‘67 Excel’ instance |
|  | \*"8 Excel" | 0 | 0 | Identified two instances of ’48 Excel’ |
|  | \*"9 Excel" |  |  | Identified a ‘29 Excel’ instance |
| app-10-19 Excel | 1# Excel | 0 | 0 | ’19 Excel’ already identified |
| app-20-29 Excel | 2# Excel | 0 | 0 | ’29 Excel’ already identified |
| app-30-39 Excel | 3# Excel | 0 | 0 |  |
| app-40-49 Excel | 4# Excel | 0 | 0 | ’44 Excel’ and two ’48 Excel’ already identified |
| app-50-59 Excel | 5# Excel | 0 | 0 | Identified a ’50 Excel’ |
| app-60-69 Excel | 6# Excel | 0 | 0 | ’67 Excel’ already identified |
| app-70-79 Excel | 7# Excel | 0 | 0 | ’67 Excel’ already identified |
| app-80-89 Excel | 8# Excel | 0 | 0 |  |
| app-90-99 Excel | 9# Excel | 0 | 0 |  |
| App-100+ Excel | 1## Excel | 0 | 0 |  |
| app-19 Excel | None | 1 | 16 | Identified in \*"6 Excel" search (a 16 Excel and a 3 Excel in same entry) |
| app-29 Excel | None | 1 | 29 | Identified in \*"9 Excel" search |
| App-44 Excel | None | 1 | 44 | Identified in \*"4 Excel" search |
| app-48 Excel | None | 2 | 96 | Identified in \*"8 Excel" search |
| App-50 Excel | None | 1 | 50 | Identified in 5# Excel search |
| app-67 Excel | None | 1 | 67 | Identified in \*"7 Excel" search |
| **Total Excel docs** |  | **215** | **625** | **Sum of all the above** |
| app-HTML | None | 119 | 382 | Calculated by adding the results of the searches on HTML documents below |
| app-1 HTML | \*"1 HTML" | 96 | 96 |  |
| app-2 HTML | \*"2 HTML" | 9 | 18 | Identified a ’12 HTML’ and a ’22 HTML’ |
| app-3 HTML | \*"3 HTML" | 3 | 9 | Identified a ’13 HTML’ and a ’23 HTML’ |
| app-4 HTML | \*"4 HTML" | 1 | 4 |  |
| app-5 HTML | \*"5 HTML" | 2 | 10 |  |
| app-6 HTML | \*"6 HTML" | 0 | 0 | Identified an ’16 HTML’ instance |
| app-7 HTML | \*"7 HTML" | 0 | 0 |  |
| app-8 HTML | \*"8 HTML" | 0 | 0 |  |
| app-9 HTML | \*"9 HTML" | 1 | 9 |  |
| app-10 HTML | 1# HTML | 1 | 10 |  |
| app-100+ HTML | 1## HTML | 0 | 0 | Identified a ‘140 HTML’ instance |
| app-12 HTML | None | 1 | 12 | Identified in \*"2 HTML" search |
| app-13 HTML | None | 1 | 13 | Identified in \*"3 HTML" search |
| app-16 HTML | None | 1 | 16 | Identified in \*"6 HTML" search |
| app22 HTML | None | 1 | 22 | Identified in \*"2 HTML" search |
| app23 HTML | None | 1 | 23 | Identified in \*"3 HTML" search |
| app 140 HTML | None | 1 | 140 | Identified in 1## HTML search |
| **Total HTML docs** |  | **119** | **382** | **Sum of all the above** |
| app-Help | None | 4 | 90 | Calculated by adding the results of the searches on Help documents below |
| app-n Help | \*"Help" | 0 | 0 | Identified all instances below |
| app-1 Help | None | 2 | 2 |  |
| app-2 Help | None | 1 | 2 |  |
| app-86 Help | None | 1 | 86 | Identified in \*"6 Help" search |
| **Total Help docs** |  | **4** | **90** | **Sum of all the above** |
| app-Zip | None | 46 | 92 | Calculated by adding the results of the searches on Zip folders below |
| app-1 Zip | \*"1 Zip" | 34 | 34 |  |
| app-2 Zip | \*"2 Zip" | 7 | 14 | Identified one ’22 ZIP’ instance |
| app-3 Zip | \*"3 Zip" | 0 | 0 |  |
| app-4 Zip | \*"4 Zip" | 1 | 4 |  |
| app-5 Zip | \*"5 Zip" | 2 | 10 |  |
| app-6 Zip | \*"6 Zip" | 0 | 0 |  |
| app-7 Zip | \*"7 Zip" | 0 | 0 |  |
| app-8 Zip | \*"8 Zip" | 1 | 8 |  |
| app-9 Zip | \*"9 Zip" | 0 | 0 |  |
| App-22 Zip | None | 1 | 22 | Identified in \*"2 Zip" search |
| **Total Zip docs** |  | **46** | **92** | **Sum of all the above** |
| app-n MPG | \*"mpg" | 20 | 20 | Identified 20 instances of ‘1 mpg’ |
| app-n fmpro7 | \*"fmpro" | 14 | 36 |  |
| app-n visio | \*"visio" | 8 | 13 |  |
| app-n ithink | \*"ithink" | 2 | 2 |  |
| app-n access | \*"access" | 2 | 5 |  |
| app-n msproject | \*"msproject" | 1 | 1 |  |
| app-n screen (Lotus ScreenCam) | \*"screen" | 1 | 4 |  |
| app-n paint | \*"paint" | 2 | 3 |  |
| app-n chp | \*"chp" | 1 | 1 |  |
| app-n bmp | \*"bmp" | 1 | 2 |  |
| app-n xmp | \*"xmp" | 1 | 1 |  |
| cdde | \*"cdde" | 34 | 34 |  |
| cddf | \*"cddf" | 37 | 37 |  |
| cdro | \*"cdro" | 25 | 25 |  |
| dbl | \*"dbl" | 622 | Unknown |  |
| empty | \*"empty" | 5 | n/a |  |
| left | \*"left" | 16 | 16 | 4 Ref items identified in the search were excluded |
| multi | \*"multi" | 8606 | Unknown |  |
| n docs | None | 17205 | 28418 | Calculated by adding the results of the searches on documents below |
| 1 doc | Omit \*"docs" | 14114 | 14114 | Since the occurrence of ‘1 docs’, was not recorded, this can only be deduced by omitting instances of n docs from the total record count. However this may have resulted in an underestimate since it eliminates occurrences of ‘doc’ elsewhere in the Movements field. |
| 2 docs | \*"2 docs" then omit in turn \*"12 docs", \*"22 docs", \*"32 docs" etc. through to \*"92 docs" | 1410 | 2820 |  |
| 3 docs | \*"3 docs" then omit \*"13 docs", \*"23 docs", \*"33 docs" etc. through to \*"93 docs" | 554 | 1662 |  |
| 4 docs | \*"4 docs" then omit in turn \*"14 docs", \*"24 docs", \*"34 docs" etc. through to \*"94 docs" | 316 | 1264 |  |
| 5 docs | \*"5 docs" then omit in turn \*"15 docs", \*"25 docs", \*"35 docs" etc. through to \*"95 docs" | 184 | 920 |  |
| 6 docs | \*"6 docs" then omit in turn \*"16 docs", \*"26 docs", \*"36 docs" etc. through to \*"96 docs" | 125 | 750 |  |
| 7 docs | \*"7 docs" then omit in turn \*"17 docs", \*"27 docs", \*"37 docs" etc. through to \*"97 docs" | 88 | 616 |  |
| 8 docs | \*"8 docs" then omit \* in turn \*"18 docs", \*"28 docs", \*"38 docs" etc. through to \*"98 docs" | 73 | 584 |  |
| 9 docs | \*"9 docs" then omit in turn \*"19 docs", \*"29docs", \*"39 docs" etc. through to \*"99 docs" | 54 | 486 |  |
| 10 docs | \*"10 docs" | 41 | 410 |  |
| 11 docs | \*"11 docs" | 46 | 506 |  |
| 12 docs | \*"12 docs" | 49 | 588 |  |
| 13 docs | \*"13 docs" | 20 | 260 |  |
| 14 docs | \*"14 docs" | 14 | 196 |  |
| 15 docs | \*"15 docs" | 14 | 210 |  |
| 16 docs | \*"16 docs" | 15 | 240 |  |
| 17 docs | \*"17 docs" | 12 | 204 |  |
| 18 docs | \*"18 docs" | 8 | 144 |  |
| 19 docs | \*"19 docs" | 3 | 57 |  |
| 20 docs | \*"20 docs" | 5 | 100 |  |
| 21 docs | \*"21 docs" | 6 | 126 |  |
| 22 docs | \*"22 docs" | 8 | 176 |  |
| 23 docs | \*"23 docs" | 7 | 161 |  |
| 24 docs | \*"24 docs" | 0 | 0 |  |
| 25 docs | \*"25 docs" | 0 | 0 |  |
| 26 docs | \*"26 docs" | 0 | 0 |  |
| 27 docs | \*"27 docs" | 1 | 27 |  |
| 28 docs | \*"28 docs" | 3 | 84 |  |
| 29 docs | \*"29 docs" | 1 | 29 |  |
| 30-39 docs | 3# docs |  |  | Identified 1 ’30 docs’, 2 ’33 docs’, 2 ’35 docs’, 1 ’36 docs’, 1 38 docs |
| 30 docs | None | 1 | 30 | Identified in 3# docs search |
| 33 docs | None | 2 | 66 | Identified in 3# docs search |
| 35 docs | None | 2 | 70 | Identified in 3# docs search |
| 36 docs | None | 1 | 36 | Identified in 3# docs search |
| 38 docs | None | 1 | 38 | Identified in 3# docs search |
| 40-49 docs | 4# docs |  |  | Identified 1 ’41 docs’, 1 ’43 docs’, 1 ’45 docs’, 1 ’46 docs’, 2 ’47 docs’, 1 ’48 docs’, 2 ’49 docs’ |
| 41 docs | None | 1 | 41 | Identified in 4# docs search |
| 43 docs | None | 1 | 43 | Identified in 4# docs search |
| 45 docs | None | 1 | 45 | Identified in 4# docs search |
| 46 docs | None | 1 | 46 | Identified in 4# docs search |
| 47 docs | None | 2 | 94 | Identified in 4# docs search |
| 48 docs | None | 1 | 48 | Identified in 4# docs search |
| 49 docs | None | 2 | 98 | Identified in 4# docs search |
| 50-59 docs | 5# docs |  |  | Identified 1 ’50 docs’, 5 ’51 docs, 3 ’52 docs’, 3 ’53 docs, 1 ’54 docs’, 1 ’55 docs’, 1 ’57 docs’ |
| 50 docs | None | 1 | 50 | Identified in 5# docs search |
| 51 docs | None | 5 | 255 | Identified in 5# docs search |
| 52 docs | None | 3 | 156 | Identified in 5# docs search |
| 53 docs | None | 3 | 159 | Identified in 5# docs search |
| 54 docs | None | 1 | 54 | Identified in 5# docs search |
| 55 docs | None | 1 | 55 | Identified in 5# docs search |
| 57 docs | None | 1 | 57 | Identified in 5# docs search |
| 60-69 docs | 6# docs | 0 | 0 |  |
| 70-79 docs | 7# docs |  |  | Identified 1 ’70 docs’, 1 ’78 docs’ |
| 70 docs | None | 1 | 70 | Identified in 7# docs search |
| 78 docs | None | 1 | 78 | Identified in 7# docs search |
| 80-89 docs | 8# docs | 0 | 0 |  |
| 90-99 docs | 9# docs |  |  | Identified 1 ’95 docs’ |
| 95 docs | None | 1 | 95 | Identified in 9# docs search |
| **Total n docs** |  | **17205** | **28418** | **Sum of all the above** |
| ord | \*"ord" then omit \*"word" | 4 | Unknown |  |
| phys | \*"phys" | 335 | Unknown |  |
| poor | \*"poor" | 53 | Unknown |  |
| pro | \*"pro" then omit \*"fmpro" | 91 | n/a |  |
| refn | \*"refn" | 4 | n/a |  |
| sold | \*"sold" | 3 | n/a |  |
| Str | \*"str" | 25 | Unknown |  |
| ***XX Problems*** | xx( | 455 | n/a |  |
| apperr-yyy xn | \*"apperr" | 162 | 296 | The # of instances was calculated from all the entries below |
| Apperr-Word | \*"apperr-word" | 12 | 22 |  |
| Apperr-PowerPoint | \*"apperr-PowerPoint" | 126 | 155 |  |
| Apperr-Excel | \*"apperr-excel" | 1 | 1 |  |
| Apperr-Visio | \*"apperr-visio" | 4 | 6 |  |
| Apperr-Access | \*"apperr-Access" | 2 | 5 |  |
| Apperr-Help | \*"apperr-Help" | 2 | 87 |  |
| Apperr-pdf | \*"apperr-pdf" | 2 | 2 |  |
| Apperr-tif | \*"apperr-tif" | 1 | 1 |  |
| Apperr-bmp | \*"apperr-bmp" | 1 | 2 |  |
| Apperr-paint | \*"apperr-paint" | 1 | 1 |  |
| Apperr-xml | \*"apperr-xml" | 1 | 1 |  |
| Apperr-html | \*"apperr-html" | 3 | 4 |  |
| Apperr-msproject | \*"apperr-msproject" | 1 | 1 |  |
| Apperr-screen | \*"apperr-screen" | 1 | 4 |  |
| Apperr-ithink | \*"apperr-ithink" | 1 | 1 |  |
| Apperr-exe | \*"apperr-exe" | 3 | 3 |  |
| cont | \*"cont" | 6 | 6 |  |
| de | \*"de" then omit \*"cdde" | 1 | n/a | Only one of 6 hits from this search was relevant |
| disc | \*"disc" | 6 | 7 |  |
| emiss-xxx | \*"emiss" | 3 | 22 | Instances calculated from instances of each emiss type |
| emiss-tif | \*"emiss-tif" | 2 | 18 |  |
| emiss-excel | \*"emiss-excel" | 1 | 4 |  |
| **Total emiss** |  | **3** | **22** | **Sum of all the above** |
| ind | \*"ind" | 0 | n/a | Issues with these 10 items were resolved after the initial check had been completed |
| lost | \*"lost" then omit OK | 278 | Unknown |  |

**References**

Wilson P A, 20 years in the life of a long-term empirical personal electronic filing study, Behaviour & Information Technology, 2001, Vol. 20, No. 5, 395-409

Wilson P A, Preservation planning for personal digital collections, a Case Note published by the Digital Preservation Coalition on its website, 11Apr2016, http://www.dpconline.org/advice/case-notes/1641-case-note-prese-planning-ers-collections-pwilson

Wilson, P.A., Checking Working Document, Unpublished document, 30 May 2016