Euan’s Save-As Corpus

They are fairly simple tables. The first one describes the applications. The piece of work was done in order to document rendering experiments we have been conducting, so the application table references the test operating system. It also includes the Pronom ID that matches most closely to the application. That particular column is not very well populated however as I gave up after seeing how few applications were documented in Pronom.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| [**idNo**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60idNo%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**name**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60name%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**version**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60version%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**testOperatingSystem**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60testOperatingSystem%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**defaultSaveAsParam**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60defaultSaveAsParam%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**defaultOpenParam**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60defaultOpenParam%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**PUIDNearestMatch**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&sql_query=SELECT+%2A+FROM+%60application%60+ORDER+BY+%60application%60.%60PUIDNearestMatch%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7)Nearest match to a PUID from PRONOM |
| 6 | Ashton Tate DBASE IV | Verison IV | MS-DOS 6.22 | dBase IV (\*.DBF) | dBase IV (\*.DBF) | x-sfw/134 |

The next four tables are identically structured and contain, respectively: the export, import, save-as and open parameters associated with each application. these were documented by opening the application and visually checking. For the most part they are complete for each application. Some of the MS-DOS applications do not have displayed parameters for import or open (although they functionally do it) and many applications do not have export p

|  |  |
| --- | --- |
| [**appID**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appExportParam&sql_query=SELECT+%2A+FROM+%60appExportParam%60+ORDER+BY+%60appExportParam%60.%60appID%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**exportParamName**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appExportParam&sql_query=SELECT+%2A+FROM+%60appExportParam%60+ORDER+BY+%60appExportParam%60.%60exportParamName%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) |
| [6](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&pos=0&sql_query=SELECT+%2A+FROM+%60enivronmentDatabase%60.%60application%60+WHERE+%60idNo%60+%3D+6&token=b695b1b0b81f651f9789842d7f80bfa7) | dBASE II (\*.db2) |

|  |  |
| --- | --- |
| [**appID**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appImportParam&sql_query=SELECT+%2A+FROM+%60appImportParam%60+ORDER+BY+%60appImportParam%60.%60appID%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**importParamName**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appImportParam&sql_query=SELECT+%2A+FROM+%60appImportParam%60+ORDER+BY+%60appImportParam%60.%60importParamName%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) |
| [7](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&pos=0&sql_query=SELECT+%2A+FROM+%60enivronmentDatabase%60.%60application%60+WHERE+%60idNo%60+%3D+7&token=b695b1b0b81f651f9789842d7f80bfa7) | dBASE II (\*.db2) |

|  |  |
| --- | --- |
| [**appID**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appSaveAsParam&sql_query=SELECT+%2A+FROM+%60appSaveAsParam%60+ORDER+BY+%60appSaveAsParam%60.%60appID%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**saveAsParamName**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appSaveAsParam&sql_query=SELECT+%2A+FROM+%60appSaveAsParam%60+ORDER+BY+%60appSaveAsParam%60.%60saveAsParamName%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) |
| [7](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&pos=0&sql_query=SELECT+%2A+FROM+%60enivronmentDatabase%60.%60application%60+WHERE+%60idNo%60+%3D+7&token=b695b1b0b81f651f9789842d7f80bfa7) | BMP - Windows Bitmap (\*.bmp) |

|  |  |
| --- | --- |
| [**appID**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appOpenParam&sql_query=SELECT+%2A+FROM+%60appOpenParam%60+ORDER+BY+%60appOpenParam%60.%60appID%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**openParamName**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=appOpenParam&sql_query=SELECT+%2A+FROM+%60appOpenParam%60+ORDER+BY+%60appOpenParam%60.%60openParamName%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) |
| [7](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=application&pos=0&sql_query=SELECT+%2A+FROM+%60enivronmentDatabase%60.%60application%60+WHERE+%60idNo%60+%3D+7&token=b695b1b0b81f651f9789842d7f80bfa7) | AutoCAD (\*.dxf) |

There is one last relevant table that documents every combination of file extension and parameter. For example if a save-as parameter for a particular application is "WordPerfect 6.0/6.1 (\*.wpd;\*.wpt;\*.doc;\*.wp)" then there are 4 file extensions associated with that parameter and so this table will have 4 entries for that parameter (one for each extension). Other applications function's parameters may share these sets of combinations so that is why they are stored in a separate table (so multiple functions across multiple applications can share the data in this table, e.g. export for app1, import for app 2 save-as for app1).

|  |  |
| --- | --- |
| [**ext**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=extParam&sql_query=SELECT+%2A+FROM+%60extParam%60+ORDER+BY+%60extParam%60.%60ext%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) | [**paramName**](http://10.1.10.5/phpmyadmin/sql.php?db=enivronmentDatabase&table=extParam&sql_query=SELECT+%2A+FROM+%60extParam%60+ORDER+BY+%60extParam%60.%60paramName%60+ASC&token=b695b1b0b81f651f9789842d7f80bfa7) |
|  |  |
| .123 | WK3;FM3 (1-2-3) (\*.wk3) |

There are a number of uses I have identified for this data so far.

1.     The work the team at the university of Freiburg have been doing on migration by emulation (e.g. <http://www.openplanetsfoundation.org/blogs/2011-09-01-word-perfect-6-rtf-migration-if-service>) could make use of this data to automate the process.

2.     The data can also be searched to find applications that open files based on extensions and parameter descriptions (if these can be associated with format signatures). It is also interesting to see whether applications agree about the need to differentiate between formats or not, e.g. if many office products from different vendors all differentiate between a version of .doc files then they are probably functionally different. On the other hand, if they all agree they all lump a number of versions under one parameter than you can assume that there is no functional difference between them (it has been pointed out that these parameter difference may be due to trying to sell the product as having compatibility with different formats even though they use the same functionality to open multiple variants/versions).

3.     The default open and save-as parameters are also useful for trying to ascertain the right environment to associate with a file for rendering in an emulator. More often than not, files that strictly adhere to a format will have been created by the application that creates that format by default, rather than one that can create them but creates files with a different format by default. If the application that creates a format by default also opens the format by default then its probably the best candidate for use in automatically rendering the file in an emulated environment.

I hope this helps evaluate the value of the data. We intend on documenting more applications as and where we get time here. We have quite a backlog. For example we recently received a large box of apple software that we want to document when we get the chance.