THE NATIONAL ARCHIVES

DROID: User Guide

OGL

Contents

July 2024

1   Introduction

1.1 What is DROID?

DROID (**D**igital **R**ecord **O**bject **ID**entification) is a file format identification tool developed by The National Archives. It is in widespread use across the world, in cultural memory institutions, local and central government departments and other public bodies, and has been embedded into multiple commercial and open source digital preservation products.

The core function of DROID is accurate file format identification, even if the file extension is wrong or missing. Where possible, identifications are made beyond the broad type, down to the version level e.g. Acrobat PDF v.1.6 - Portable Document Format. DROID can currently identify over 1400 file formats, and this number is growing all the time. Information about file formats including the identification signatures utilised by DROID are kept within PRONOM, The National Archives' file format registry.

In addition to identifying the file format DROID also extracts other information about the files it scans, such as, file size, last modified date and file path. This information is presented in a profile which can be analysed on screen in the DROID Graphical User Interface (GUI) using filtering, or exported to a CSV file. DROID also examines the files within container files, such as 'zip' files, as well as the container file itself.

This guidance is based on the current version of DROID (v.6.6), while some of the advice may also be applicable to earlier versions of DROID (or indeed later ones), certain specific features may not be available, or may behave differently.

1.2 What is the purpose of this guidance?

This guidance will help you to:

- understand how to install and run DROID across your digital files

- interpret your results, avoiding common pitfalls

- understand some of the potential drivers for profiling your files with DROID

1.3 Who is this guidance for?

This guidance is aimed at knowledge, information and records management professionals, archivists, or digital preservation practitioners who want to gather information about digital files in their care. Note: some of this guidance goes into technical detail, so users may want to seek assistance from their IT support, for instance with help in installing DROID. IT professionals may also find it helpful to refer to this guide, but will also find 'Help' files on hand within DROID.

2   Running DROID

2.1 Installing and using DROID

DROID is an open source tool, available for free download. You can download the latest version [here](#) as a zip file. From DROID 6.5 onwards each release provides two download options for users, the first is for Windows only and contains embedded Java and the second for any system with Java 8 to17 or OpenJDK installed. To install DROID unzip the files into a suitable folder on your computer.

If using the second option DROID requires Java 8 to 17 of OpenJDK and should work on any platform which supports this.

To run DROID go to the folder where it is installed. If you are using Microsoft Windows then double-click on the file called 'droid.bat'. If you cannot see the droid.bat file check that the 'Hide extensions for known file types' box is unticked in your folder options. If you are running Apple Mac or Linux double-click on the file called 'droid.sh'.

These instructions refer to the GUI version of DROID; if you require instruction for using DROID in the command line please see the help section in the DROID GUI.

If you are experiencing issues running DROID, first try the following:

- If using the version of DROID without embedded Java, check which version of Java is installed (by typing java –version in CMD) and install version 8 to 17 if required

- Check DROID is running from a location on the local machine and not on a shared network/drive

- Close down DROID and delete the .droid6 folder (within Windows, usually found here: C:\Users\'*username*'\.droid6) then restart DROID which will re-create the .droid6 folder

- If using a Linux or OSX operating system you may have to set executable permissions to be allowed to run DROID. You should set the executable permission on the 'droid.sh' file. You may also need to set executable permissions on the two .jar files: 'droid-ui-6.8.0.jar' and 'droid-command-line-6.8.0.jar'

If you have any further problems installing or running DROID, email us at: PRONOM@nationalarchives.gov.uk.

2.2 'About DROID' dialog

When the DROID GUI is started, you can navigate to the menu item Help -> About which launches a dialog with information about the current DROID build, Java runtime version in use, its location as well as the location of the .droid6 and log folders.
There is a copy button on this dialog which copies all this information onto clipboard so that it can be easily pasted into an email when asking for help from Droid technical support. This information can be useful for troubleshooting common problems or finding out what version of java is on your machine.

July 2024
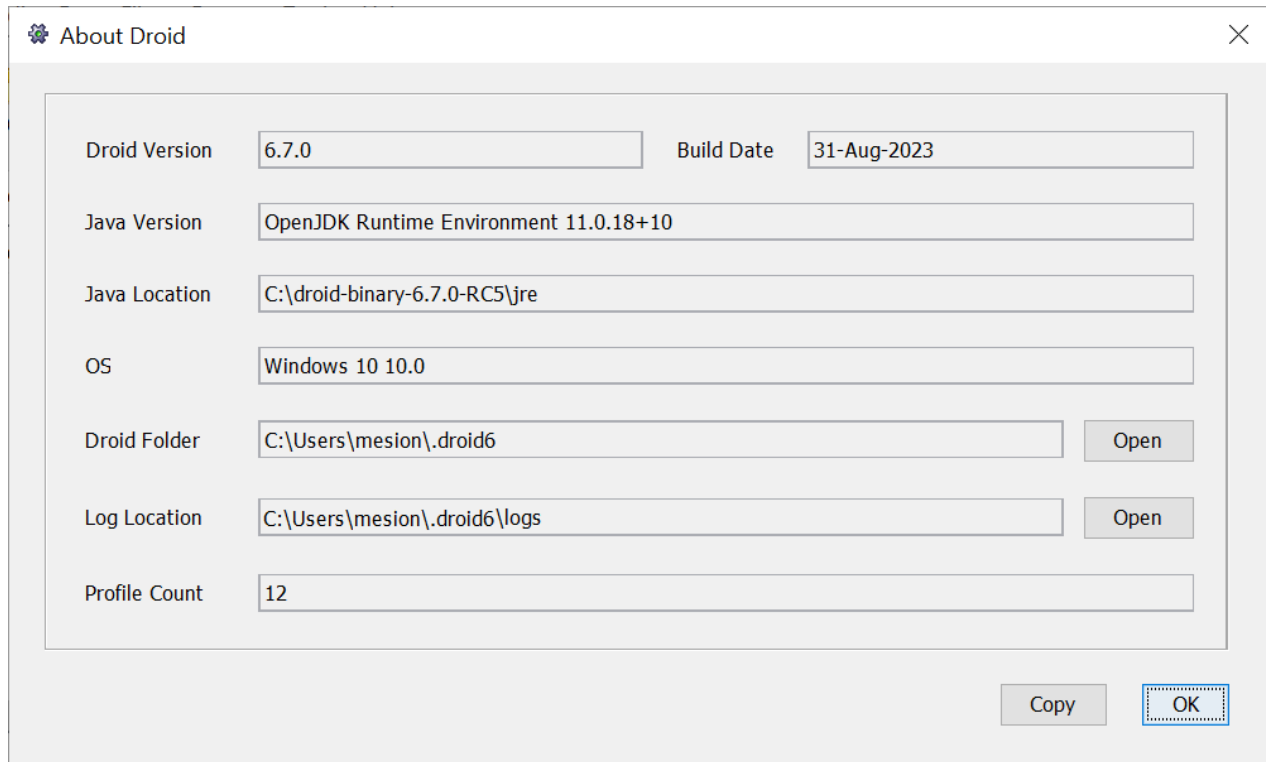
2.3 Creating a profile

DROID will automatically create a new profile for you when it is first started (a blank tab called Untitled-1). This can sometimes take a little while to fully initialise but once it is ready it should appear as the screenshot does below. You are able to add files to a profile while it is initialising but you will not be able to start running DROID over them until it is ready.
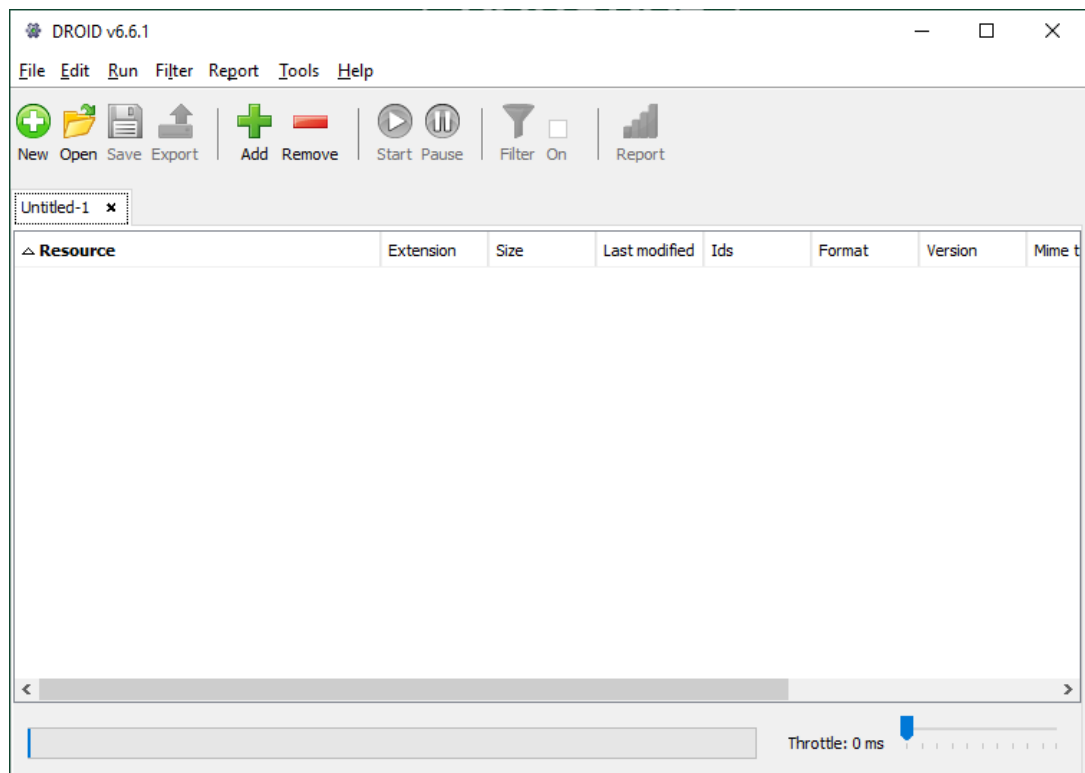
You can create multiple profiles and run them at the same time. Each profile will appear in its own tab underneath the toolbar. The profiles can be renamed. Once a profile is created its settings are fixed. See the Appendix for the different settings preferences available.

2.4 Choosing files and folders
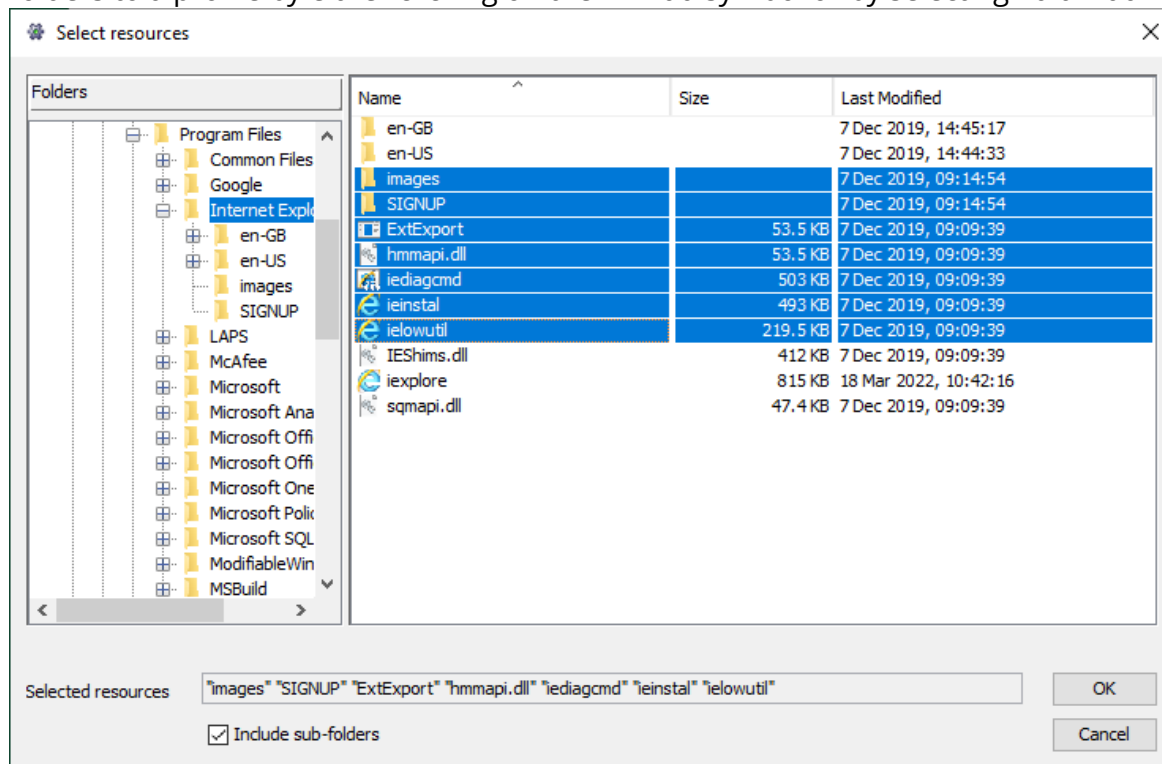
Files and folders can be dragged and dropped into the GUI. Alternatively, you can add files and folders to a profile by either clicking on the ✚ Add symbol or by selecting Edit>Add Files/Folders.

On the left hand side is a navigator, which allows you to explore your file system. The example above shows the 'Internet Explorer' folder selected on the left hand side, with the contents of this folder shown on the right. The box 'Include subfolders' is ticked by default. Uncheck this box if you only want to profile the files directly inside a folder and ignore any sub-folders.

If you accidentally add a file or folder that you don't want to profile, you can simply remove it. Note: you can't add or remove files or folders after a profile has been started – only while you are specifying what you want to profile.

2.5 Running a profile

Once you have added all the files or folders you wish to profile you can then run the profile. Start by pressing the Start button ▶. The progress bar will display while the profile is running; this will give you an approximate idea of progress. This is calculated by the amount of files being scanned, it does not account for the size of the files or for files which exist within other files (e.g. zip files).
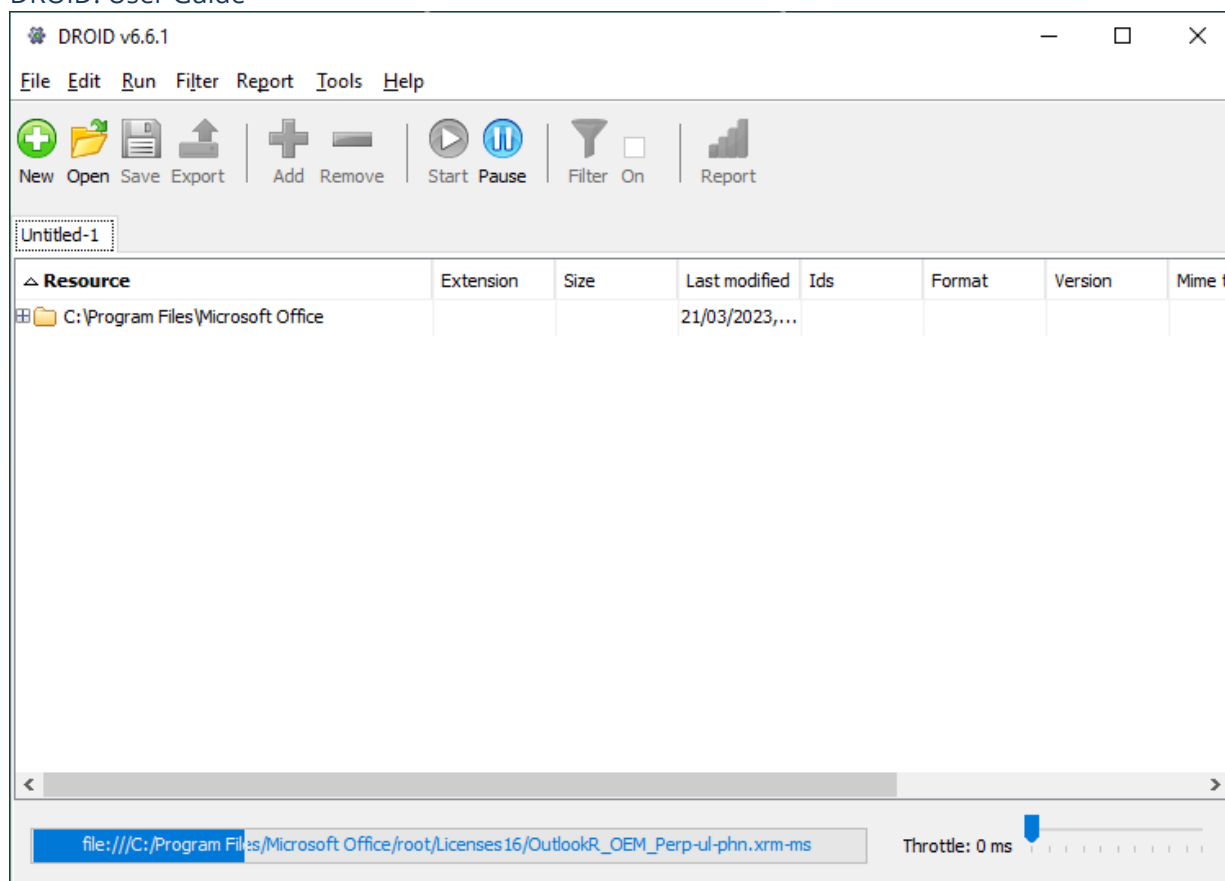
If required you can reduce the amount of load DROID is using on your computer, for example, if it is slowing down other applications on your computer. The throttle bar can be moved, allowing you to adjust the speed accordingly. By default it is set to run as fast as possible.

You can pause a profile at any time by pressing the pause button. Note: it may take a little while for the pause to take effect, as DROID queues up work which runs in parallel. It must wait for all these tasks to complete before it can pause.

When a profile is paused, you can save, filter, export and create reports on the results generated so far. If a profile is paused, you can resume it by simply re-pressing the start button.

2.6 Saving a profile
You can save the profile by pressing the save button. This will save as a '.droid' file.[1] This allows you to reload your profile at any time for analysis or to resume a paused profile.

2.7 Exporting to CSV
One of the most common ways to analyse your results is to export them as a CSV and use a spreadsheet application to view them. The examples shown in this section use MS Excel but other spreadsheet applications can be used as well.

To export your results press the Export button. This should open the Export profiles window.

---

[1] Droid files are actually zip files, which contain some XML files describing the profile and a database containing any results of profiling so far. DROID currently uses the Apache Derby database, version 10.7, which can be opened using various third-party tools, such as DB Visualizer. The username to connect to a droid database is **droid_user**, and the password is the same as the username.

July 2024

The profiles you have open are listed on the left side under 'Select profiles for export'. If a profile is empty or in the process of running it will be greyed out. Select all the profiles you want to export by checking the boxes next to them, n.b. if any of your profiles have active filters, then the results will also be filtered.

When the "Use export template" checkbox is unchecked, you have the ability to select specific columns to be included using checkboxes next to the column names on the right.
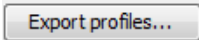

When the "Use export template" checkbox is checked, you have the ability to choose one of the pre-configured export templates. The profiles are exported as per the columns and their modifications defined in the export template.

July 2024

If you wish for the contents of each column to be enclosed in quotes then there is a tick box 'Quote all exported columns'. This could be useful so that a delimiter, if used as a value in a field can be treated correctly by other software.

You can also select whether the export should produce one row per file, or one row per format. When exporting one row per file, each row in the CSV file will represent a single file, folder or archival file profiled with DROID. If exporting one row per format, each row in the CSV file will be a single format identification made by DROID. Since a file can be identified as being more than one possible format, this option will produce CSV files with multiple rows for the same file (but with different identifications for it).

When you are ready, press the Export profiles... button. This will bring up a standard file-save box allowing you to specify where you would like to save your CSV file. Make sure that you select to save the file as a comma separated values file (*.csv).
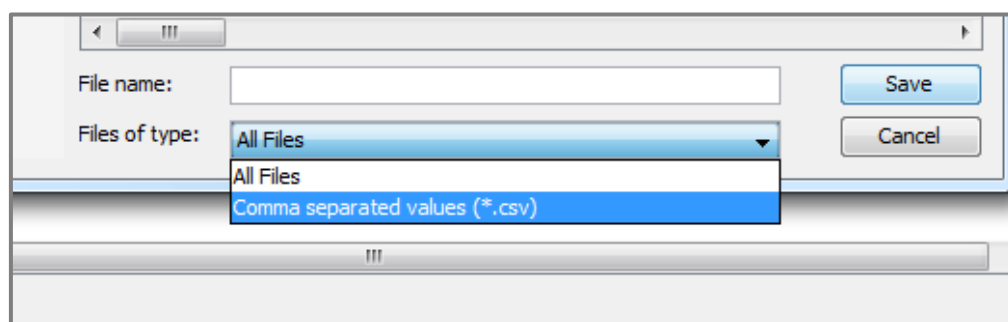
If you are surveying large amounts of files and wish to use the CSV export you should be aware of spreadsheet application limits. The limits of a workbook for Excel are 1,048,576 rows by 16,384 columns. Be aware that when using functions such as conditional formatting and filtering, these can slow down considerably when surveying a large number of cells. In order to avoid this you can split your DROID profiles into more manageable sizes and then survey the smaller CSVs separately.

3   What information does DROID gather?

DROID extracts metadata about your files and folders. A CSV export will contain the following columns and you will see the main sub-set of these on screen within the GUI:

1. ID (A unique number assigned to each file, folder or archival file)

2. PARENT_ID (ID of the archival file or folder in which the file is contained)

3. URI (Uniform resource identifier)

4. FILE_PATH (File system location for this resource)

5. NAME (File name for this resource)

6. METHOD (Identification method, whether it is "signature", "container signature" or "extension")

July 2024

7. STATUS (Identification status of the resource

8. SIZE (file size in bytes)

9. TYPE (Type of the resource, whether it is a file, folder or container)

10. EXT (The file extension of resource)

11. LAST_MODIFIED (Date and time when this resource was last modified)

12. EXTENSION_MISMATCH (Whether there is a mismatch between the extension of the resource and its identification by signature)

13. HASH (Fixed size string of character generated by running the contents of the file through one of the hashing algorithms. Droid supports MD5, SHA1, SHA256 and SHA512)

14. FORMAT_COUNT (Number of identifications made for this resource)

15. PUID (PRONOM unique identifier for the file format)

16. MIME_TYPE (Mime type of this resource)

17. FORMAT_NAME (The name of an identified file format name)

18. FORMAT_VERSION (The version number of an identified file format)

The following sections describe each of the above items of information and where any issues can arise in understanding and interpreting them.

3.1 ID and parent ID
DROID gives each file and folder a consecutive number in the ID column. By looking at the parent ID column you can see which files have the same parent folder.

3.2 URI and file path
DROID records the location of every file and folder it scans. It records location within the file system in two ways: using a Uniform Resource Indicator (URI), and the file path. DROID treats file paths and URIs as case sensitive.

For example, if we have a:

1. folder called 'Folder' on the 'C:\' drive of a Windows computer

2. file called 'Document.doc' inside 'Folder',

3. container file called 'Archive.zip' inside 'Folder'

4. spreadsheet called 'Spreadsheet.xls' inside 'Archive.zip'

5. folder called 'Another folder' inside 'Archive.zip'

6. picture called 'Large picture.jpg' inside 'Another folder'

Then we have the following file paths and URIs:

File path	URI
`C:\Folder`	`file:/C:/Folder/`
`C:\Folder\Docume nt.doc`	`file:/C:/Folder/Document.doc`
`C:\Folder\Archiv e.zip`	`file:/C:/Folder/Archive.zip`
	`zip:/file:/C:/Folder/Archive.zip!Spreadsheet.xls`
	`zip:/file:/C:/Folder/Archive.zip!Another%20folder/`
	`zip:/file:/C:/Folder/Archive.zip!Another%20folder/L arge%20picture.jpg`

Only files, container files or folders which are directly accessible in the file system have a file path. Files and folders which are inside the zip file do not have a file path, but do have a URI. This tells you that they are inside the zip file, where they can be found in it, and where the zip file they are inside is to be found.

The prefixes of a URI tell you what sort of resource is being described by the URI, and the exclamation marks indicate where one type of resource is contained by another. For example, for 'Spreadsheet.xls', we can see that there is a file 'C:/Folder/Archive.zip' with the prefix 'file:/'. The exclamation mark (!) tells us that the spreadsheet is contained by the 'Archive.zip' file, and the first prefix 'zip:/' tells us the type of the containment is a zip file. Note: spaces in URIs are encoded by '%20', and folder separators are always forward slashes.

URIs and file paths can be used as unique identifiers for digital files and by copying and pasting the file path into windows explorer you can easily pull up a specific file for review.

3.3 Filename
The name of a file, folder or container file is its name independent of its location within a file system or inside a container file. It includes any file extension as part of its name. DROID treats all filenames as case-sensitive. For example, 'MYDOCUMENT.DOC' and 'mydocument.doc' are regarded as different file names.

3.4 Identification method
DROID has three different methods of identifying file formats:

- extension

- signature

- container

An 'extension' identification means that a format was identified purely on the basis of its file extension. Such an identification may not be reliable, as files can be named in any way, and many file formats and versions of file formats have the same extension.

A 'signature' identification means that a format was identified by finding a specific pattern in the byte sequence, usually in the header of the file, which is unique to a particular file format and version. This method is much more reliable than identification by extension only.

A 'container' identification means that a format was identified by finding embedded files, often with

signatures of their own, inside the main file. For example, OpenDocument word processing files are actually zip files containing xml files, images or other resources used in the document. A container identification would identify the main file as an OpenDocument file, not a zip file. This method is very reliable, as not only does the broad type of container have to be identified (e.g. zip), but the zip file must then be opened and files inside scanned for further identifications to be made.

3.5 Status

As DROID scans your files and folders it records whether the profiling was successful or not. There are four different statuses which a file or folder can have (as outlined below):

Done	Success, DROID could read the file or folder, and no errors occurred while trying to do so. It does not necessarily mean that DROID could identify the file format.
Not found	The resource was moved or deleted before it could be scanned.
Access denied	The operating system refused read access to DROID. You may want to change the permissions on those resources to enable profiling.
Error	An error occurred while trying to read the file. Anything that prevents DROID from profiling a file or folder (other than being not found or access denied) will result in an error status. You may be able to determine the cause of the error by examining DROID's error logs.

3.6 File size

The size of a file or container file is recorded in bytes. Files can have a size of zero (no content, just a record in the file system). Folder size is not reported by DROID. The size of a container file is the size of the container file itself, not the sum of the sizes of its contents. For example, zip files compress their contents, so the sum of the sizes of the files inside a zip file will be bigger than the size of the container file itself.

3.7 Type

- file

- folder

- container

Format identification is made against files. Folders do not have any format identifications, sizes, or hashes reported against them but they can contain other folders, files and container files inside them. Container files are like folders, in that they can contain other folders, files and container files inside them, but they are also files themselves, so DROID will provide format identifications, file sizes and hashes for them. Presently, DROID can look inside zip, tar, gzip, rar, 7zip, iso, bzip2, arc and warc container files.

3.8 File extension

File extensions are readily recognisable and indicate the broad type of a file (or container file) by appending a short string to a file name, separated by a full stop e.g. .doc, .pdf, .xls. DROID converts

all file extensions it extracts to lower case, to facilitate sorting, reporting and filtering. Note: DROID will only extract an extension for files, not folders, and there must be at least one character before the full stop i.e. there must be a name before the extension.

3.9 Last modified date

Most files, folders and container files record the date and time when they were last modified. This is not the same as the date a file was originally created or the date when a file was last read. It is possible that not every file, folder or container file will have a last modified date. For example, in some cases resources inside container files may not record this date.

It is important to note that last-modified dates can be changed when files are copied from one server to another, so this date may not reflect the last date a user actively modified the content of a file. Also, the content of a file (the data within it) may actually be older than the file itself – if a file was copied, or simply typed up manually from an older piece of content.

Some files may have noticeably inaccurate dates, e.g. 1 Jan 1970. In this case, the files will be newer than indicated. This error will likely be caused by the battery failing on the internal clock of the computer from which the document was uploaded.

Despite this the last-modified dates are often the most accurate dates that can be extracted in an automated way from a file system.

DROID reports last modified dates in the ISO date time format: YYYY-MM-DDTHH:MM:SS

3.10 File extension mismatch warning

Sometimes file extensions are incorrect or missing. If DROID detects that the file extension for a file name does not match the format of the file it has identified by signature or container signature, it will issue a file extension mismatch warning. For example, if a file called 'myfile.doc' is identified as a spreadsheet, then a file extension mismatch warning will be issued. File extension mismatches can alert you to the fact that users may have trouble opening a file by double clicking on it.

3.11 Hash (checksum)

DROID can optionally generate a content hash or checksum of the contents of each file and container file, using either the SHA 1, SHA 2 (256) or MD5 algorithms. A content hash is a long unique string of numbers and letters that can be used to uniquely identify the content of the file. It is extremely unlikely that two different files will have the same content hash. Content hashes can be used to detect files with duplicate content, or can be linked to forensic hash databases to find or exclude files which are widely used (and therefore not unique to your organisation) or which contain illegal content. Hashes can also be helpful when moving files around. Once moved, if you re-calculate the hashes and check against the previous values you can verify that the copying process has not caused any changes at the bit level.

Detecting duplicates

An easy way to see if you have duplicate files using Excel is to highlight the HASH column and select the Home section - Conditional Formatting - Highlight Cells Rules - Duplicate Values. You can then choose a colour which Excel will use to highlight each cell which has duplicate values.
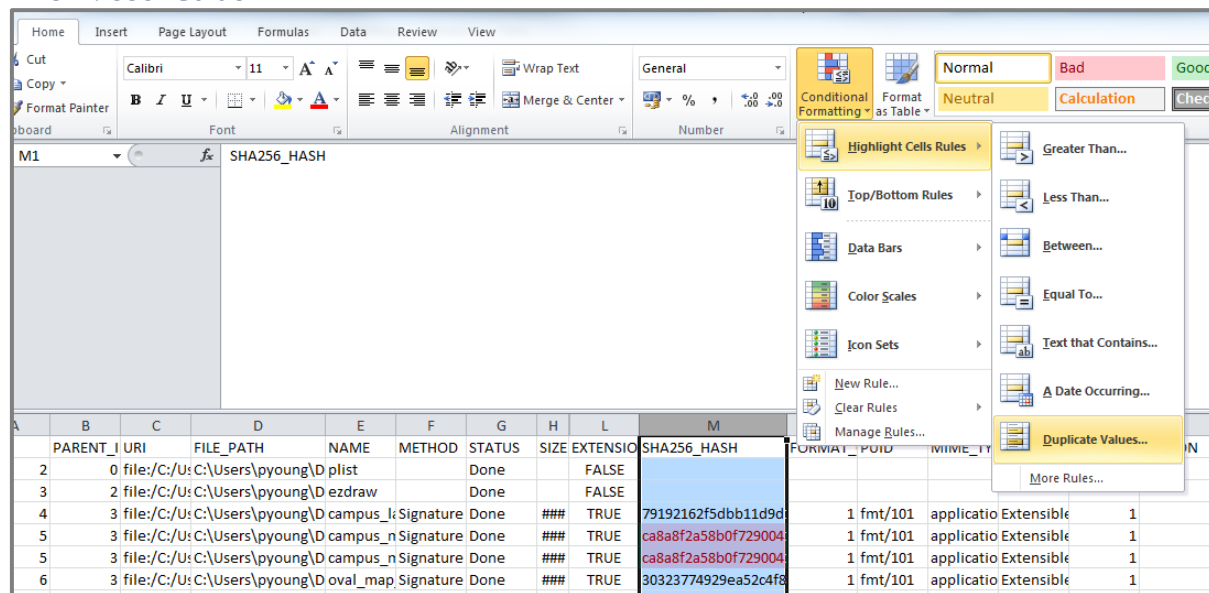
3.12 File format count

DROID attempts to identify the format of files, including container files, but not folders. The number of identifications DROID records for a file can vary. It can have:

- zero, if DROID can't identify a format at all

- one, if it is a clear match

- more than one, if it matches multiple extensions or identification signatures

The latter situation can happen for two reasons:

1. A format is identified purely on the basis of its file extension; multiple versions of a file format often have the same extension.

2. A format has several versions and variations which are very similar and something needs to be tweaked in the signatures for these formats or in the priorities assigned to them to make a definitive identification.

If you have files which are not identifying or have multiple identifications you can potentially examine why this is happening yourself by looking at the signature recorded for the file format you were expecting the files to identify as, within the PRONOM registry. If you examine your files in a Hex Editor you will be able to see where the file differs from the expected signature. The National Archives is invested in improving file format identification signatures and helping users with file format identification issues, so users are encouraged to contact us via the PRONOM submission form or at PRONOM@nationalarchives.gov.uk.

Please include as much information as possible regarding null or multiple identifications, including if they are possible sample copies of the files themselves.

3.13 PRONOM Unique Identifier (PUID)

PRONOM assigns unique identifiers to each format added to the registry, these are effectively consecutive numbers prefixed by 'fmt/' or 'x-fmt/'. Clicking on a PUID in the DROID GUI will take you

to the relevant page for that file format on The National Archives' PRONOM website, where you can access information about the file format. PUIDs are guaranteed never to change, although new PUIDs will be defined.

3.14 Mime-type

Mime-types, now more commonly referred to as media-types, are two part identifiers used to identify file formats in use on, and transmitted via, the Web. They are assigned by a body called the Internet Assigned Numbers Authority and take the form of a type and sub type separated by a forward slash e.g. image/bmp. Mime-types are quite broad classifications, so many different file formats may have the same mime-type. For example, the mime-type for PUID 'fmt/40' is 'application/msword' – which is shared by all other Microsoft Word formats.

3.15 Format name

As well as the PUID and mime-type DROID will report a file format name when it identifies a file format. The names given to file formats in PRONOM and reported by DROID aim to balance accuracy as well as being reflective of the common way a file format is referred to in practice e.g. JP2 (JPEG 2000 part 1).

3.16 Format version

In addition to the PUID, mime-type and format name, DROID will also report on which version of a particular file format a file has identified as, as long as this distinction exists within the PRONOM registry. The National Archives endeavours to keep PRONOM up to date with new versions of file formats that already exist within the registry and likewise aims to add all versions of a file format when adding a new format to PRONOM. As new versions are released over time this may result in a generic PUID for a file format and additional version specific PUIDs. Identification accuracy is maintained by assigning higher priority to specific versions over the generic format. As with file format names, versions reported by DROID reflect the way versions are termed by software proprietors or how they are commonly referred to – this may mean the version is referenced within the format name as well as separately reported by DROID e.g. 'Acrobat PDF 1.2 - Portable Document Format' is the format name and '1.2' will additionally appear in the version column.

Note on on-screen exploration

The DROID GUI offers a platform to navigate and explore your results on screen, in the same way as file managers on Windows, Mac or Linux. Your scanned files are displayed as files and folders, nested within one another. To open a folder or archival file click on the plus sign beside it. To close it again click on the minus sign. You can sort the information on screen by clicking on the column headers.

If a file has no identifications then the file will have a grey dot in the Ids column. If a file has a single identification then the file will have a blue cog in the Ids column, and format information will appear in the format, version and mime-type columns. If a file has multiple identifications then the format columns will be blank, and a number in brackets will appear in the Ids column, indicating the number of different identifications made. Clicking on this number will bring up a small window containing all the identifications made. Note: file extension mismatch warnings are displayed in the extension column as a yellow warning triangle icon.

If you want to look at the files themselves on your computer, DROID provides a way to open the nearest folder containing the selected file or files. Select the 'Edit/Open Containing Folder' or right-

click and select 'Open Containing Folder'. Your normal file manager will open at the folder which contains the item. If you select a file inside an archival file then you will be taken to the folder containing the archival file.

4 Filtering in the GUI

If you have scanned a large number of files, it can be time consuming to manually look for files which are of interest to you. If you apply a filter to a profile, then all the files and folders which don't meet your criteria are filtered out, leaving only the ones that you are interested in.

4.1 Defining a filter

Click on the filter button at the top of DROID to bring up the filter box, as shown below. Filters consist of one or more filter criteria (rules for filtering). Each filter criterion consists of three things: Field, Operation and Values.

For example, you could define a filter like this: Field: File size; Operation: Greater than; Values: 1000000, as in the screenshot below:
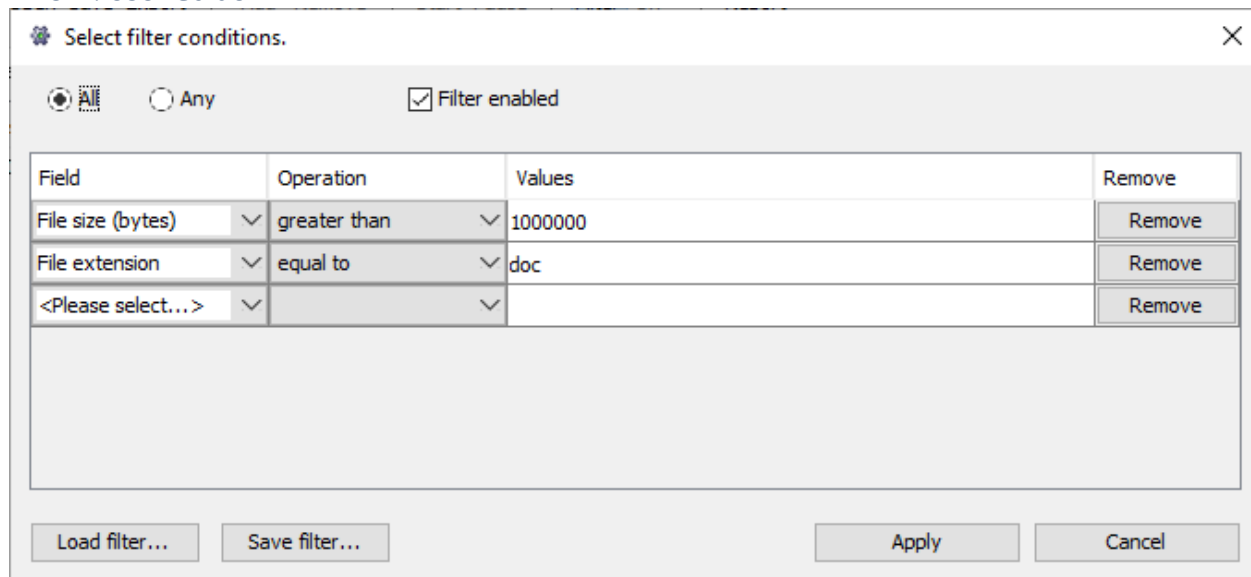

Also, if you only wanted to look at files with the extension 'doc', you could add a second filter criterion:

July 2024

You can add as many filter criteria as you like to pinpoint the files or folders you are interested in. In the above example, you are looking for files which meet all of the criteria: files of a size bigger than 100000 bytes and which have a file extension equal to 'doc'. You can also look for files which meet any of the criteria. For example, you may define a filter like this:
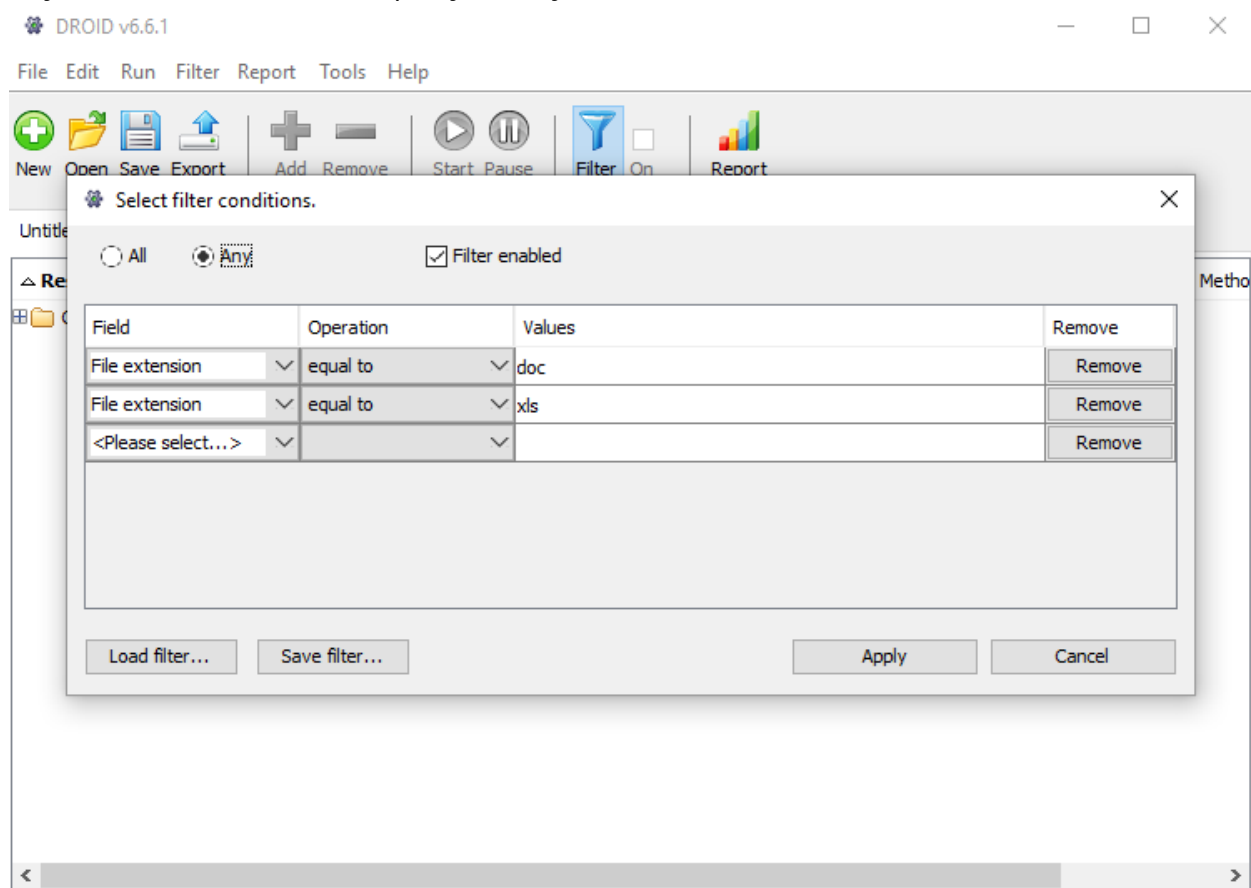
This filter would locate files which have any of the file extensions. Whether a filter looks for any or all of the criteria is set by simply switching the any/all option on the filter editing window.

4.2 Managing filters

Each profile has its own filter, independent of other profiles, allowing you to define different filters for different profiles. You can save and load filters you use frequently, so you don't have to keep

manually creating them. You can also copy a filter to all open profiles, if you want them to have the same filter on each.

If you define and enable a filter it applies to what you see on screen, any reports you generate, and any exports you perform. This allows you to produce reports and exports over only the items of interest. Filters can also be turned on or off quickly using a button in the GUI, as shown below.
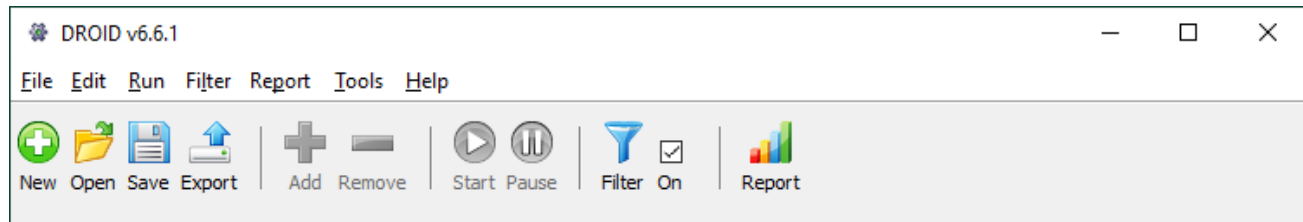
Note: there is one important difference when filtering items on screen. Any folders or archival files which are needed to let you navigate to the unfiltered items are also displayed (or else you could never get to the unfiltered items). These items are displayed with greyed-out icons, without any information other than their name – this indicates that they do not meet your filter conditions, but are needed for navigational purposes. They will not appear in any reports or exports you perform.

5 Reporting via the GUI

Exploring and filtering your profiles allows you to see the kinds of files you have. However, if you want broad statistics over all your files, then DROID offers a variety of quick reports that let you see the bigger picture. If you have enabled a filter then the report will be on the filtered items only, letting you get different statistics for different sets of files and folders. Reports offer the following statistics (where possible):

- count of items

- total size of items

- minimum size of items

- maximum size of items

- average size of items

Note: the count of items may include folders as well as files (depending on the report you are running). Also, be aware that the count is of all the files scanned, which may include files inside zip files or other archival files. Hence, the   count may be higher than a count of files provided by your operating system. If you have excluded processing archival files then the count should be the same as that reported by your operating system.

You can report on one or more profiles at the same time. If you are reporting on more than one profile then you will be given statistics for each profile and the totals across all profiles.

5.1 Breakdowns
Some reports offer these statistics broken down by another value. For example, DROID can produce statistics broken down by the year files or folders were last modified.

One important feature to note is: if your report is broken down by file formats (file format PUID or MIME type) then you will probably find that the final total count and sizes of the files are bigger than you will see in other reports. This is because files can be identified as more than one potential file format. Hence, when breaking down by file format, the same file can appear against more than one file format in the statistics. When adding up the totals for each breakdown a file can be double-counted in the final totals. This is not incorrect, as the report is producing totals per breakdown, then totalling across all the breakdowns. However, you should be aware of this when interpreting the results.

5.2 Included reports
The reports included with DROID are:

- total numbers and sizes of files and folders

- total unreadable files

- total unreadable folders

- count and size of files by the year last modified

- count and size of files by the month last modified

- count and size of files by the year and month last modified

- count and size of files by the file format PUID

- count and size of files by the file extension

- count and size of files by the MIME type

- comprehensive breakdown (all of the above in a single report)

**6		Appendix: Configuring DROID**

6.1 Setting Preferences
You can adjust the default settings of DROID to suit your needs. If you go to Tools and then Preferences you can see a range of settings which can be adjusted from their defaults.

6.2. Profile Defaults
The first tab in the Preferences box is called 'Profile Defaults'. This section allows you to change the way each profile will use DROID to scan files or directories.

Tick box 1: Analyse contents of archive files
If this option is checked then DROID will look inside zip, tar and gzip files and profile the files it finds inside them.

Tick box 2: Analyse contents of web archive files
If this option is checked then DROID will look inside arc and warc files produced by crawling web pages and profile the files it finds inside them. If the option to analyse archive files is not checked as well, DROID will only look inside unzipped arc and warc files. Note: the file type stored in a web archive file will often not be the same as the file type of the web page that produced it (e.g. a GIF image generated by a PHP page).

July 2024

Tick box 3: Generate hash for each file

If this option is checked then DROID will produce a hash (checksum) for the content of the file. By default this option is off, as generating hashes slows down profiling significantly. The hash type defaults to MD5, but SHA256 and SHA1 can be selected as alternatives.

Maximum bytes to scan

This setting allows you to configure how far from the beginning or end of the file DROID will scan before it stops trying to identify the file. This can improve the performance of DROID but could have an effect on the accuracy of its identifications.

Almost all byte patterns which identify the format of files are found fairly close to the start or end of the file. By default, this setting is 65536 bytes (64KB). You can make it smaller – increasing DROID's performance – but the accuracy of identifications may go down. Alternatively, you can make it bigger – decreasing DROID's performance – but the identification accuracy may go up.

Setting this value to a negative number (e.g. -1), will cause DROID to scan the entire file (possibly more than once, if different patterns trigger those scans). This setting gives the maximum possible accuracy DROID can achieve, but can cause DROID to profile very slowly, particularly if you have large files.

If you do have files which are not being identified, you can increase this value, or set it to -1 to see if this has any effect on identification accuracy.

Match extensions if no other signatures defined

If this option is selected DROID will only attempt to match file extensions for formats where no other binary or container signatures are defined. The thinking here is that, if an ostensibly more accurate signature failed to match the file format, then matching against the extension alone would give potentially misleading results. This is the default setting for DROID.

Match extensions against all known extensions

If this option is selected DROID will attempt to match a file extension against all known extensions. This may be helpful if you find you have files in newer formats where there is currently no good binary or container signature, or you have unusual files where the existing signatures fail to match. Using this option will produce identifications for more files, and more multiple identifications for files.

Default throttle

This is the delay in milliseconds that DROID should pause between identifying files read from the file system. Specifying a higher delay will cause DROID to work slower, placing less load on your computer, network or disk storage. It does not cause a pause between identifying files inside archival files.

Unless you need to slow DROID down, this should be set to zero. Unlike the other profile preferences, this value can be drastically adjusted while running DROID using the throttle slider control on the main window. The throttle setting can be different for each profile, and will be saved with each individual profile.

6.3 Export Defaults

DROID can export profiles to CSV files in two ways: one row per file, and one row per format. The settings here only configure the default for export. When you export you can select which you prefer on a per-export basis.

6.4 Export Templates

DROID can export profiles to CSV files using export templates. This option gives more power to the user to only include specific columns, change order in which the columns appear, change the header for the column or add additional columns. This can be achieved by creating an export template and copying it into the ".droid6\export_templates" folder.

Syntactically, the export template is simple text file with a .template extension, which defines customisations of columns to be exported. Using a template, you can customise the headers, add new columns to the export, convert the data to uppercase / lowercase and change the column order in the export. Details of the template definition are given below:

Version line: The very first line of the Export Template defines the version number. For now it simply needs to be version 1.0

Modifying a header: If you intend to use a different column header instead of one of the well-known column names used by Droid, you can do so by defining a line such as Identifier: $ID this instruction changes the header of the ID column to be "Identifier"

Adding a column: If you need to add a blank column to the export, it can be done by defining it with a header without any data. e.g. If you want to add a column named "ProcessDate", but the date is actually populated using some other method later in the process, you can define a line in the template as ProcessDate:

Adding a column and populating it with some value: If you need to add a column to the export and populate it with some value, it can be done by defining it with a header and the value in double quotes. e.g. If you want to add a column named "Submitter" and populate it with a value of "Librarian", you can define a line in the template as Submitter: "Librarian"

Modifying a header and data: If you intend to use same or different column header and want to modify the data that is populated from the profile results, you can do so by defining a modifier line such as: Path: LCASE($FILE_PATH) This line will export the value of FILE_PATH in lowercase. Please note, in version 1.0 the only supported modifications are conversion to uppercase or lowercase. Please also note that the LCASE or UCASE modifiers can only be used with the profile columns. These modifications are not supported on custom columns

Order of the columns: When using an Export Template, Droid exports only those columns which appear in the template and in the order how they appear in the template. This gives you the control over ordering of columns in the export file. If there is any column that get repeated due to multiple identifications (e.g. PUID) such column is appended immediately after the original.

Contents of an example export template file is shown below:

```
version 1.0
Identifier: $ID
Name: $NAME
LowerName: LCASE($NAME)
Submitter: "Government Dept."
Export Tool:
```

The potential output when using this template could appear as shown below:

Identifier	Name	LowerName	Submitter	Export Tool
2	GP-led-sites-may21.csv	gp-led-sites-may21.csv	Government Dept.	
3	Hospital-Hubs-may21.csv	hospital-hubs-may21.csv	Government Dept.	
4	Pharmacies-May21.csv	pharmacies-may21.csv	Government Dept.	
5	All-Sites.xlsx	all-sites.xlsx	Government Dept.	

6.5 Signature Updates

Every time PRONOM is updated a new signature is released. PRONOM creates two types of signature files: binary and container. These are what DROID uses to identify file formats. DROID is set to automatically prompt you to download the latest update by default. You can manually check for new updates by selecting Tools and then 'Check for signature updates'.

Alternatively, you can manually add a signature file to DROID by using the Tools/Upload Signature option. This allows you to download signatures and manually transfer them on to a machine which may not be connected to the Web e.g. for security purposes.

If you have any problems installing the latest updates, check that DROID is checking the correct locations for new signature files; these can be viewed by going to Tools - Preferences - Signature updates.

The Binary Signatures URL should be set to:
https://www.nationalarchives.gov.uk/PRONOM/service.asmx
The Container Signatures URL should be set to:
https://www.nationalarchives.gov.uk/pronom/container-signature.xml

In the signature updates area you can also select your preferred binary or container signatures, in case you wanted to compare the results of two different signatures. If you do select a new binary or container signature you will need to create a new profile before the changes take place.

Tick box 1: Automatically check for updates
If this box is checked then DROID will automatically attempt to look for updated signatures according to the schedule you specify.

Tick box 2: Prompt before downloading
If this box is checked DROID will prompt you before downloading new signatures.

Tick box 3: Use updated signature files by default
If this box is checked DROID will automatically use a newer signature file by default. If it isn't checked it will still download the file – it will be available for selection in the profile default tab, but it will not automatically make it the default signature file to use.

Update settings
Every time DROID starts up it will try to check for signature updates. Every X days on starting up DROID will check for signature updates after the number of days since the last check. Note: if you leave DROID running for more days than is specified, it will not automatically try to update its

signatures. The check is still only made on startup, but only if the number of days since the last time it checked has elapsed.

The PUID URL pattern

This is the web location which DROID will connect to when clicking on a PUID hyperlink in the profile tab. This should contain the URL to connect to, and the place into which the PUID should be inserted, indicated by %s. For example, if we were to click on a hyperlinked PUID of 'fmt/111' in the main profile results, DROID would launch your web browser at the address: 'http://www.nationalarchives.gov.uk/pronom/fmt/111'.