

Preparing born-digital records for transfer to The National Archives

Tools to use

In order to complete this process you will need the following software on your computer:

- Teracopy (or equivalent software): <http://codesector.com/downloads> using either the 2.3 or 3.5 (depending on which version of Windows you are running)
- DROID: <http://www.nationalarchives.gov.uk/information-management/manageinformation/preserving-digital-records/droid/>
- DroidConvert: <http://www.nationalarchives.gov.uk/information-management/manageinformation/preserving-digital-records/droid/> (located at the bottom of the page)
- CSV Validator: <https://digital-preservation.github.io/csv-validator/>
CSV Validator is used by The National Archives to check the integrity of metadata csv files.
- If you have problems downloading any of the software then please contact the Digital Transfer team via governmenttransfers@nationalarchives.gov.uk

There is a glossary included at the end of this document that provides information regarding some of the terminology used in these instructions.

Please note that the folder and file titles used throughout this document are examples only, using the series reference RW_33 for the purposes of these instructions.

How to create a DROID CSV export

Step one: creating a working area

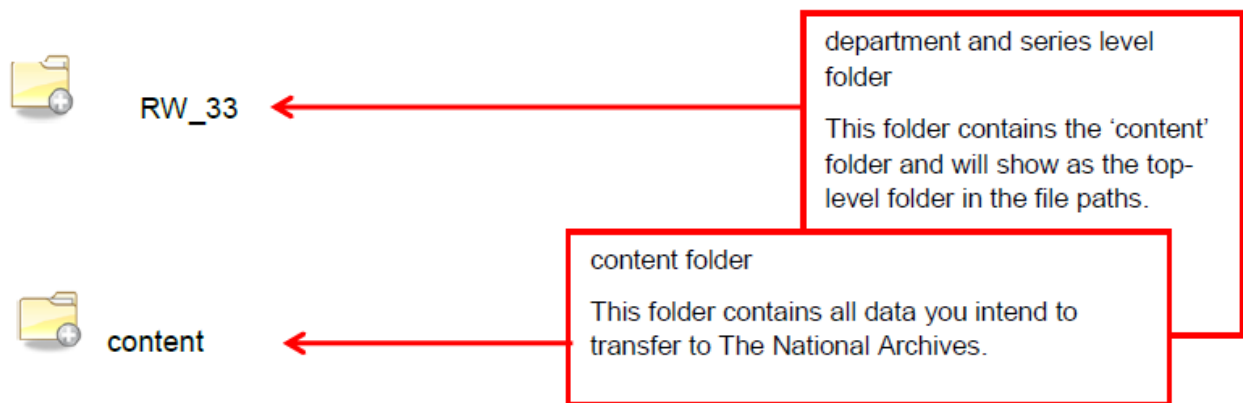
Before starting step one, please ensure:

- ✓ You have completed the appraisal section of the [Digital Transfer Form](#) up to the point where you are required to submit a DROID CSV export of the records selected;
- ✓ You have discussed with The National Archives about the method you are using to transfer your born-digital records – either via encrypted hard disk drives (HDD) or via Secure File Transfer Protocol (SFTP) if you are sending only a small volume of open records.
- ✓ If sending via encrypted HDD, The National Archives will provide an HDD encrypted using BitLocker and delivered by secure courier after you have completed validation. We will then send the encryption key by email after you have confirmed receipt of the drive. If sending via SFTP, you might find it useful to set up a separate preparation area for the following steps.

1. Create a series folder (RW_33, for example) on your encrypted HDD or preparation area using the series reference agreed with The National Archives as the folder title. If the series reference is not known at this time use the placeholder 'XX'. Within that folder, create another folder called 'content'. Ensure you create this folder with a lower case 'c'. The records will sit inside their original folder structure inside this 'content' folder, the metadata files you create will sit alongside the content folder rather than inside it;

2. Use Teracopy (or equivalent software) to copy all the folders and files that will be part of your transfer into the content folder. Teracopy ensures that the dates of the records as well as other metadata remain unchanged during the transfer process and there is no corruption in copying;
3. Once you have set this up it is vital that you **only work on your HDD or preparation area**. Where the files are located becomes important later in the process, so ensure that you run DROID and complete all the below steps on the files in the content folder on your HDD or preparation area.

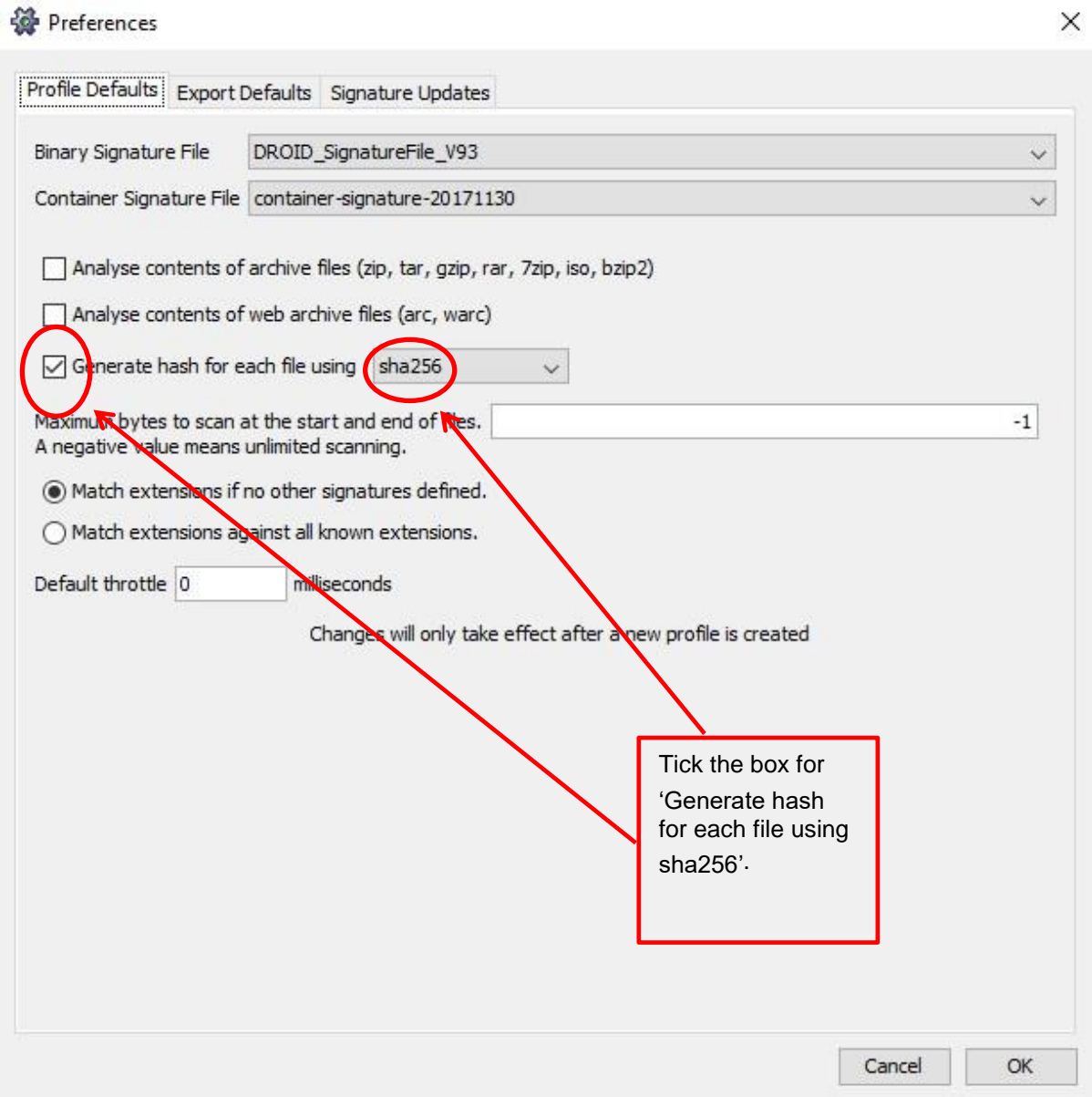
Example of folder structure on hard drive/preparation area



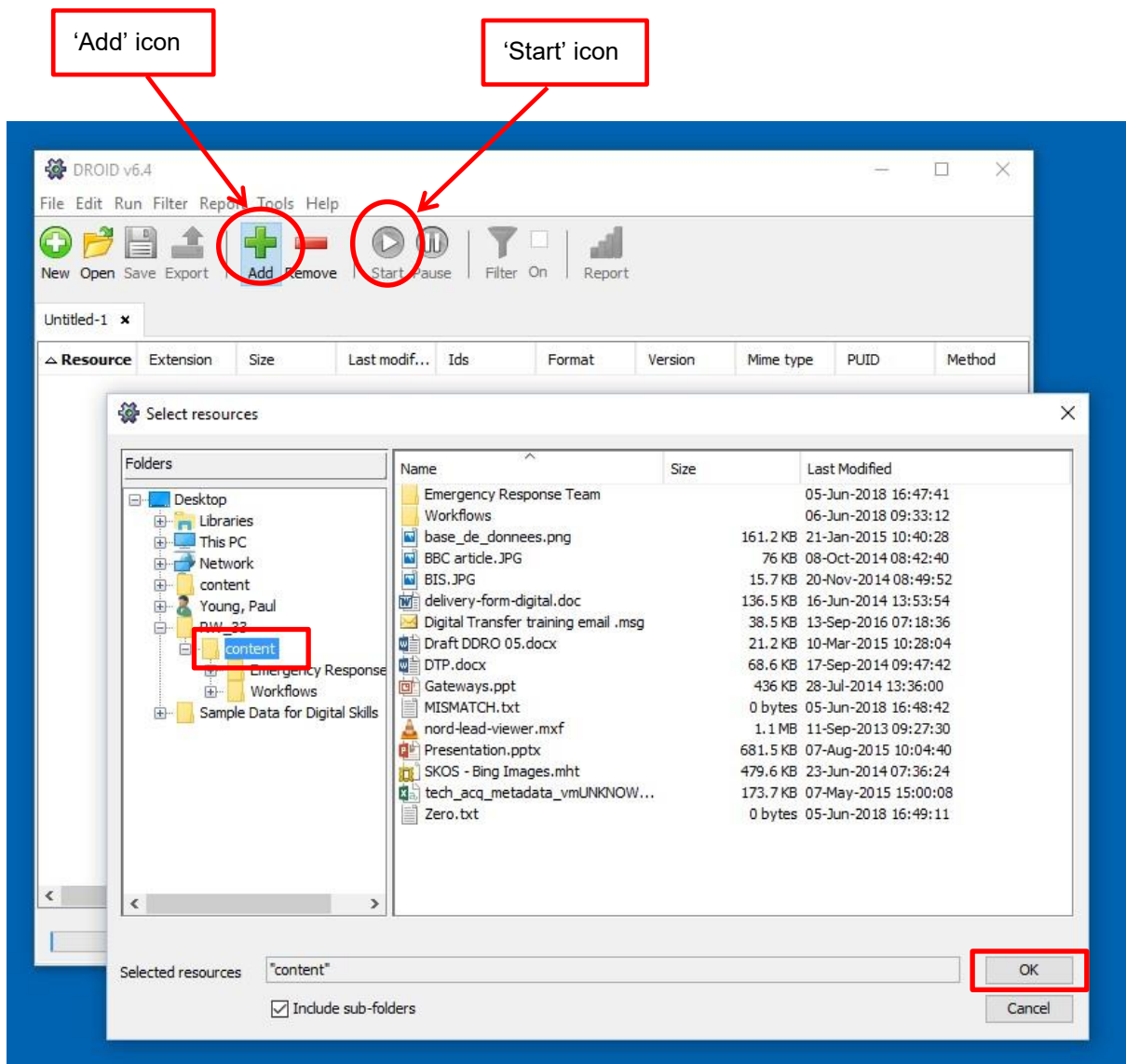
If you have more than one HDD, you will need to take all of the steps below for each HDD (this is known as a 'batch') separately. Please be aware that all metadata naming should reflect to which 'batch' it relates. So if your first HDD contains batch 1, all metadata filenames should end 'B001', likewise your second HDD is batch 2, so all metadata filenames should end 'B002' and so on.

Step two: generating an initial DROID CSV export

1. Open DROID. To run, double click on the Windows batch file 'droid.bat' (.bat extension may not display);
2. Ensure that the option to create a SHA-256 checksum is active. To do this, select Tools > Preferences and ensure that the 'Generate hash for each file using sha256' box has been ticked. If it is giving the option to create an 'md5' checksum instead of 'sha256', open the drop down box to select 'sha256'. If this option is not active yet then select it. Tick the box and then open a new profile window to ensure that this preference has been enabled;

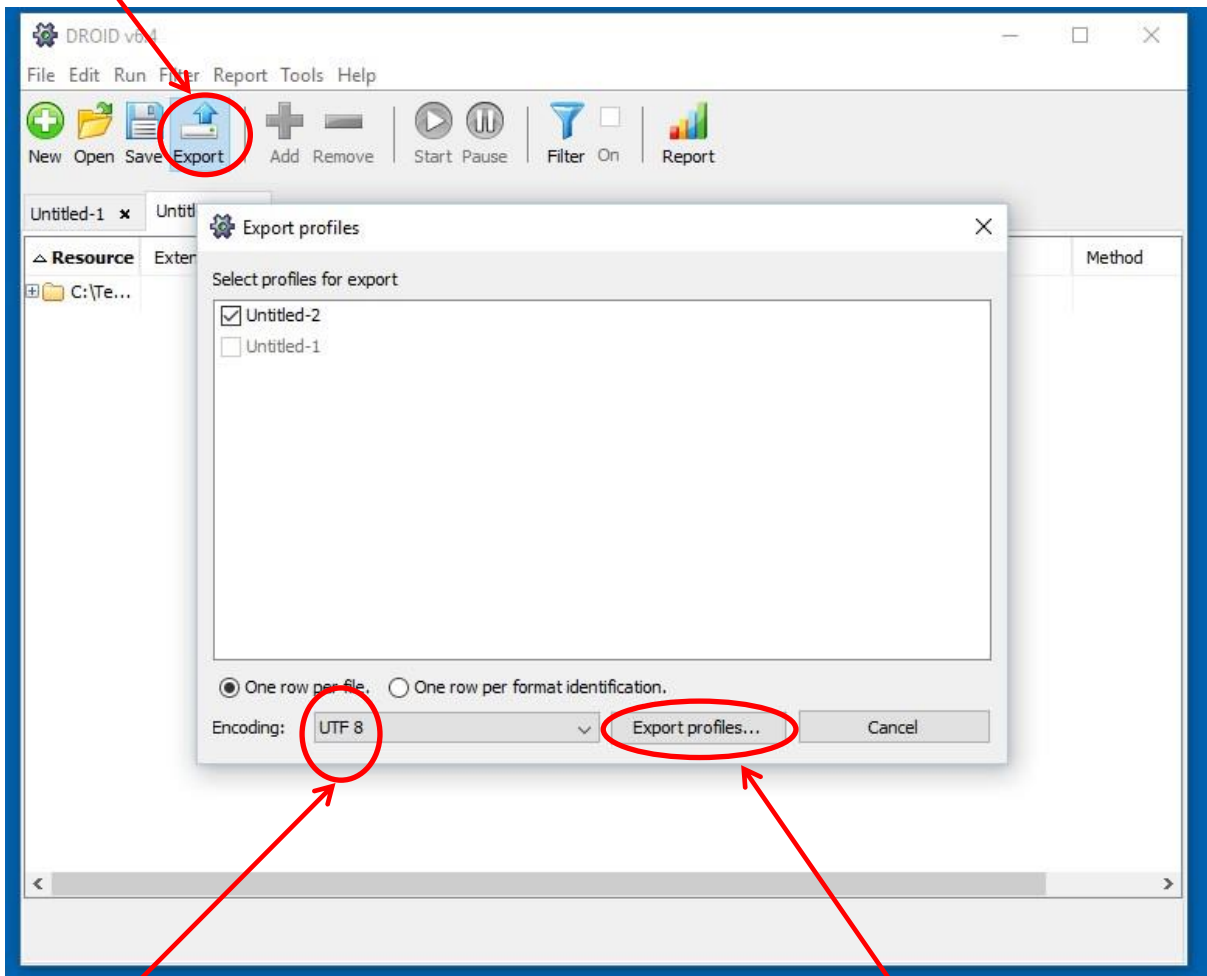


3. Select the green 'Add' icon on the main screen (as seen in the diagram below). This will open Windows Explorer. At this point, navigate to and select the 'content' folder as shown in the image below. Once you have selected the folder, it will then appear on the main DROID screen. Click 'OK';



4. Press the 'Start' icon to run DROID (this will turn blue after you click 'OK');
5. Once DROID has finished running, you will be able to export the results as a CSV file. To do this, first select the 'Export' icon and tick the box labelled 'Untitled-n' (with 'n' representing the profile you wish to export, here '-2'). Ensure the encoding at the bottom is set to UTF 8. Then, click 'Export profiles'. You will then have the option to save the export.

'Export' icon



Ensure the coding is set to UTF 8.

Export profiles in order to save the report.

- 6. Save a copy named 'DROID Export.csv'. Send a copy of the 'DROID Export.csv' to your Digital Transfer Adviser (via governmenttransfers@nationalarchives.gov.uk or direct email). This helps us with our preservation planning. It also allows us to review the data and provide feedback on any actions required prior to transfer.

Step three: generating final DROID export

Before starting step three, please ensure:

- ✓ You have received feedback on the initial DROID CSV export from your Digital Transfer Adviser.

1. Carry out any amendments required from initial feedback of DROID export (e.g. deleting zero byte files, unpacking zip files, deleting thumbs.db files);
2. If there were any amendments, run DROID over the files again using the process described in step two and create an export CSV.
3. For the purposes of these instructions, we will refer to the DROID export (either the original if unchanged, or your latest version if changes were made) as 'DROID Final.csv'. This will be the DROID export which you should use to complete your metadata csv.
4. If there are no amendments then you can move on to step four using the original 'DROID Export.csv';

How to create, generate and validate the metadata CSV files

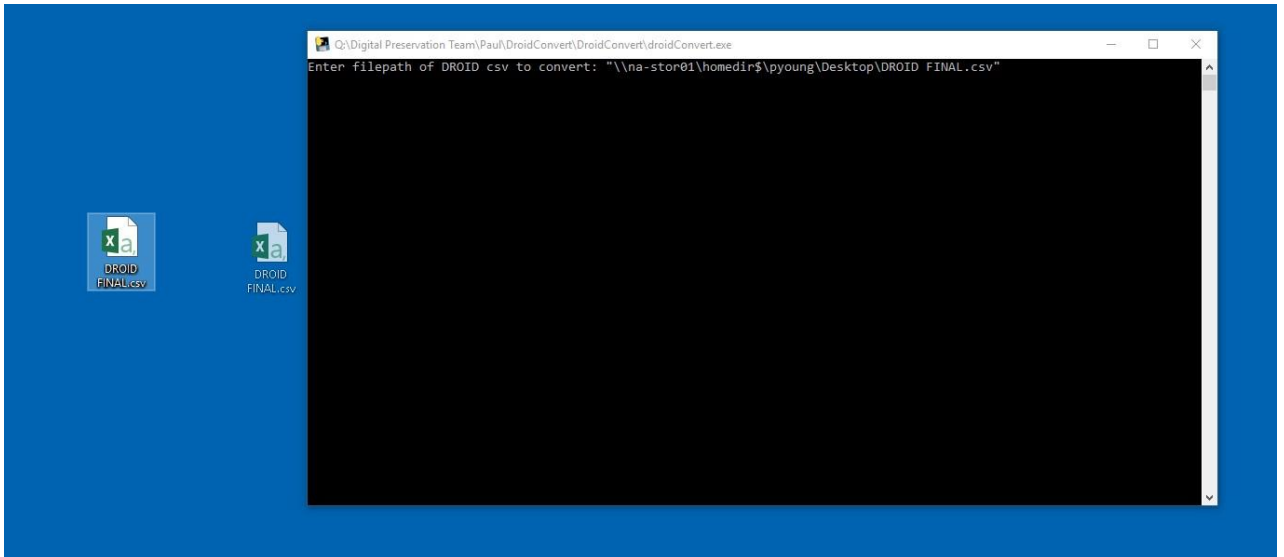
Step four: use *DroidConvert* to transform 'DROID Final.csv' to metadata CSV template

The script *DroidConvert* will take the 'DROID Final.csv' and convert it into the default fields for The National Archives metadata, and populate the standard fields.

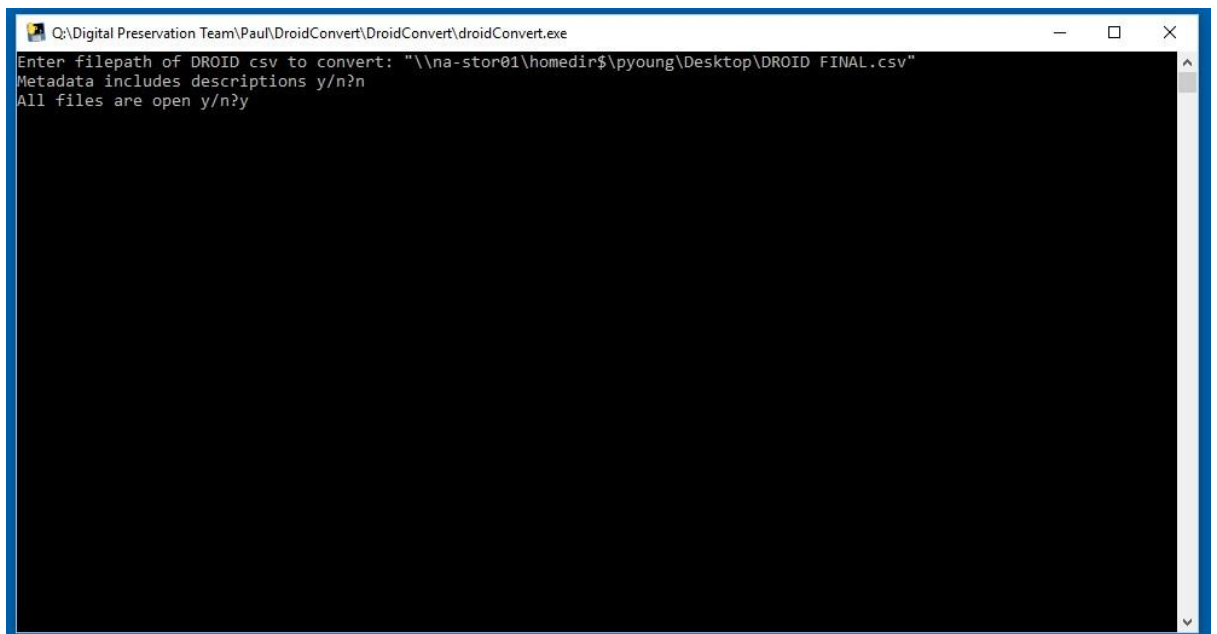
1. Double click on 'DroidConvert.bat'. This will open a window that looks like the image below;



2. Drag the 'DROID Final.csv' into the window, this will generate the filepath to the DROID final csv in the terminal window. Once added press the 'Enter' or 'Return' key;



3. The script will then ask if 'Metadata includes descriptions y/n?'. If you wish to provide descriptive metadata then enter 'y'. This will then add in the required metadata columns for descriptions. If you do not have any descriptive metadata to add in then enter 'n';
4. The script will then ask whether 'All files are open'. If all of the files in the collection are open press 'y'. If some or all of the files are closed, press 'n'. Once added press enter. If the majority of the files are open and only some are closed, you may want to choose 'y' and then change the metadata manually for the closed records;
5. Press any key to exit the script. This will create the output file, which is called 'DroidConvertoutput_DROID Final.csv'. This will be saved in the same folder that contains 'DroidConvert.bat';



6. You should now have your DROID data and all of your metadata headings in your metadata CSV template. Save this on your HDD or in your preparation area. If this is an accrual and you already have a schema file, you can name this using the same

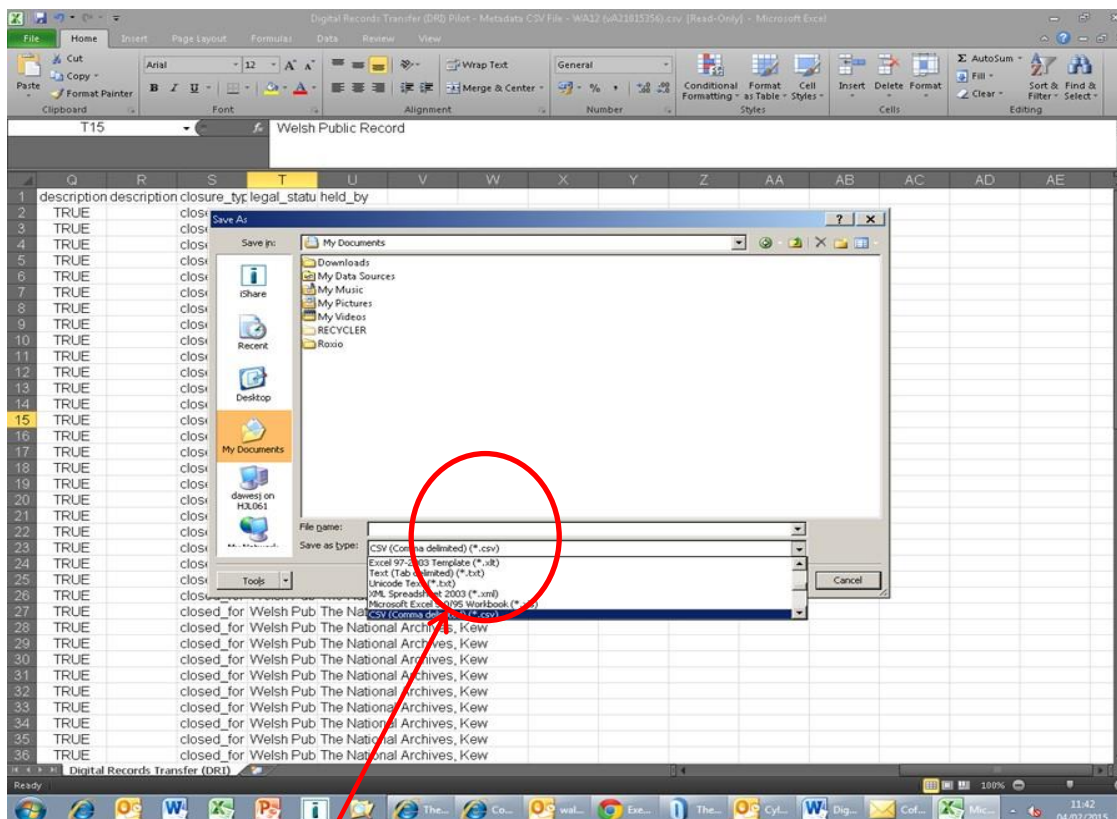
convention ('metadata_[schema version number]_[series code][Year of transfer][batch number].csv', e.g. **metadata_v16_RW33Y18HB001.csv**). If you have not received this yet you can remove the schema version number, e.g. **metadata_RW33Y18HB001.csv**. If sending multiple batches the batch number at the end of the file name will change depending on which batch you are working with. You are now ready to start filling in any remaining metadata.

Step five: completing the metadata CSV template

1. The columns 'identifier', 'file_name', 'folder', 'date_last_modified' and 'checksum' should now contain metadata generated by DROID and do not require any further changes to be made;
2. For the final three columns, DroidConvert will have populated these with the default entries, as displayed below. If your metadata requires different entries (e.g. 'Welsh Public Record'), please discuss these with your Digital Transfer Adviser;

Heading	Default Entries
rights_copyright	'Crown Copyright'
legal_status	'Public Record(s)'
held_by	'The National Archives, Kew'

3. If all of your files are open, then DroidConvert should have populated the seven columns containing closure information (or nine columns, if metadata includes descriptions). These do not require any further editing. If there are closed files then you should edit these manually using the guidance in Appendix A;
4. If your metadata contains descriptive metadata, you can add this in the 'description' column. If you have multiple columns of additional metadata you wish to add, please discuss this with your Digital Transfer Adviser;
5. If you have redacted material included within your records please contact your Digital Transfer Adviser to receive guidance for populating metadata for redacted records;



Ensure you save your completed metadata CSV file as a **CSV (comma delimited) (*.CSV file)**. This will ensure that you will be able to validate it with the CSV Validator.

6. Email your completed metadata CSV file to your Digital Transfer Adviser for checking. This will help to reduce the number of errors when you validate your metadata. If this is a new transfer, your Digital Transfer Adviser will send you a schema (.csvs) file. If this is an accruing transfer, your Digital Transfer Adviser will confirm if there are updates to your current schema.

Step six: validating your metadata

Before starting step six, please ensure:

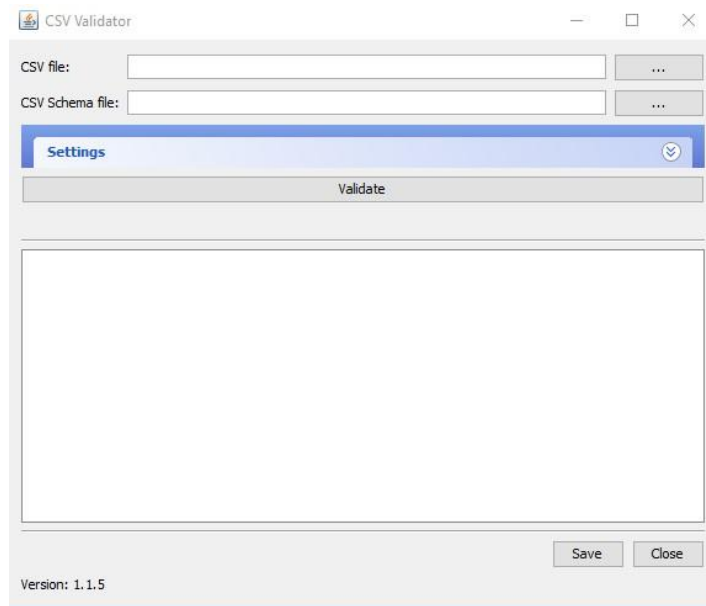
- ✓ You have completed the sensitivity review section of the [Digital Transfer Form](#) up to the point where you are required to submit an interim metadata CSV;
- ✓ You have reviewed your born-digital records and their metadata (file names and descriptions) for sensitivity. The Advisory Council on National Records and Archives has granted all closures applied for and you have completed the closure columns in the metadata CSV;
- ✓ Your Digital Transfer Adviser has provided you with a schema file (.csvs) and has provided feedback on your metadata CSV.

1. The National Archives will have provided you with a schema, in this example this is called 'metadata_v16_RW33Y18HB000.csvs'. Once you have received this, if not

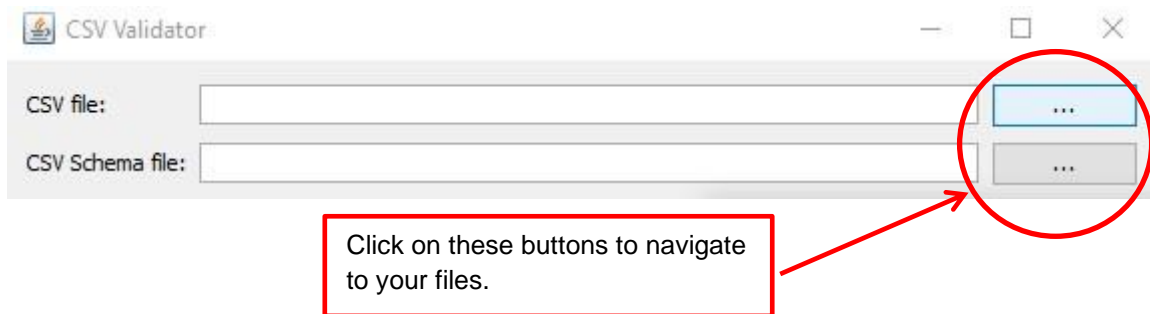
done already, you can rename the metadata file to match the filename structure, including the schema version number. In this example this would be 'metadata_v16_RW33Y18HB0001.csv';

2. The metadata schema (.csvs file) provided by your Digital Transfer Adviser should be copied into the HDD or preparation area outside of the content folder;
3. In order to run the CSV validator, open the CSV validator folder and select the 'bin' folder and then double click to run either:
 - validate-gui.bat on Windows operating systems,
 - or; □ validate-gui shell script on Unix/Linux.

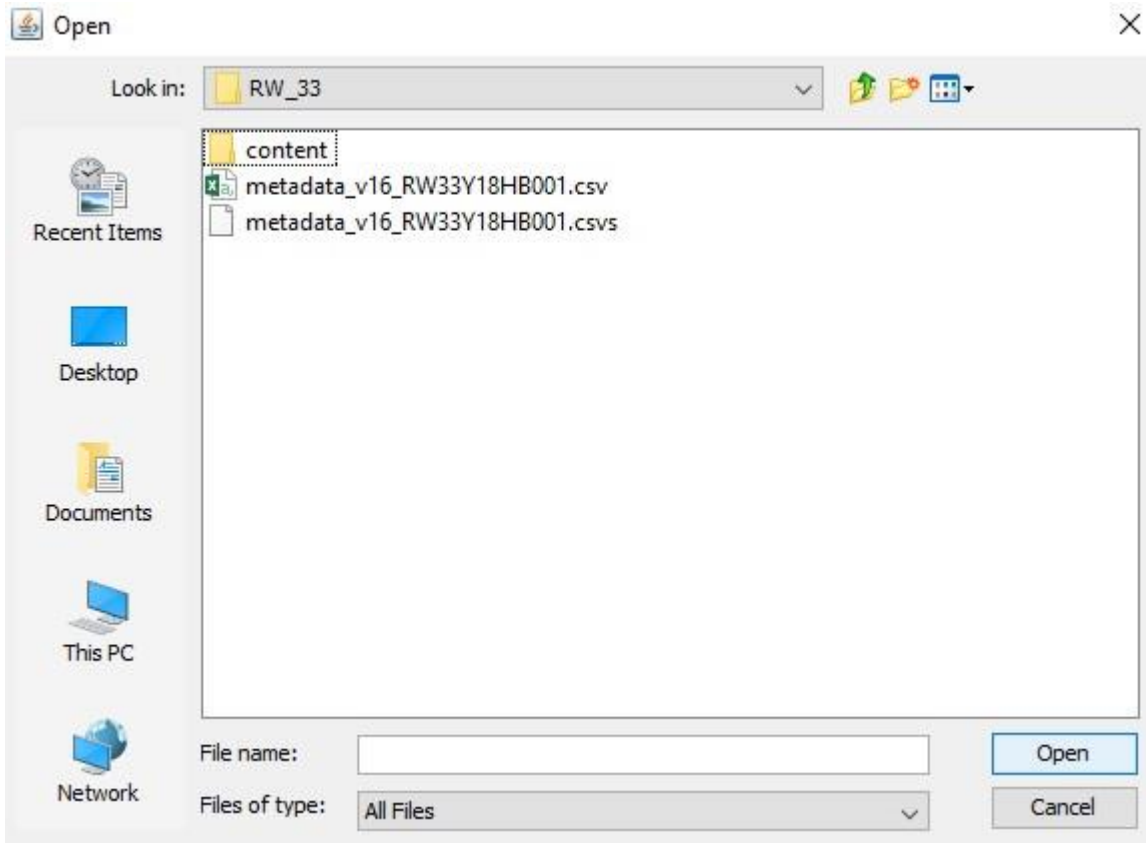
This screen should appear:



Click the '...' buttons at the top to navigate to the location of your metadata.csv file and your schema file on the HDD or preparation area.

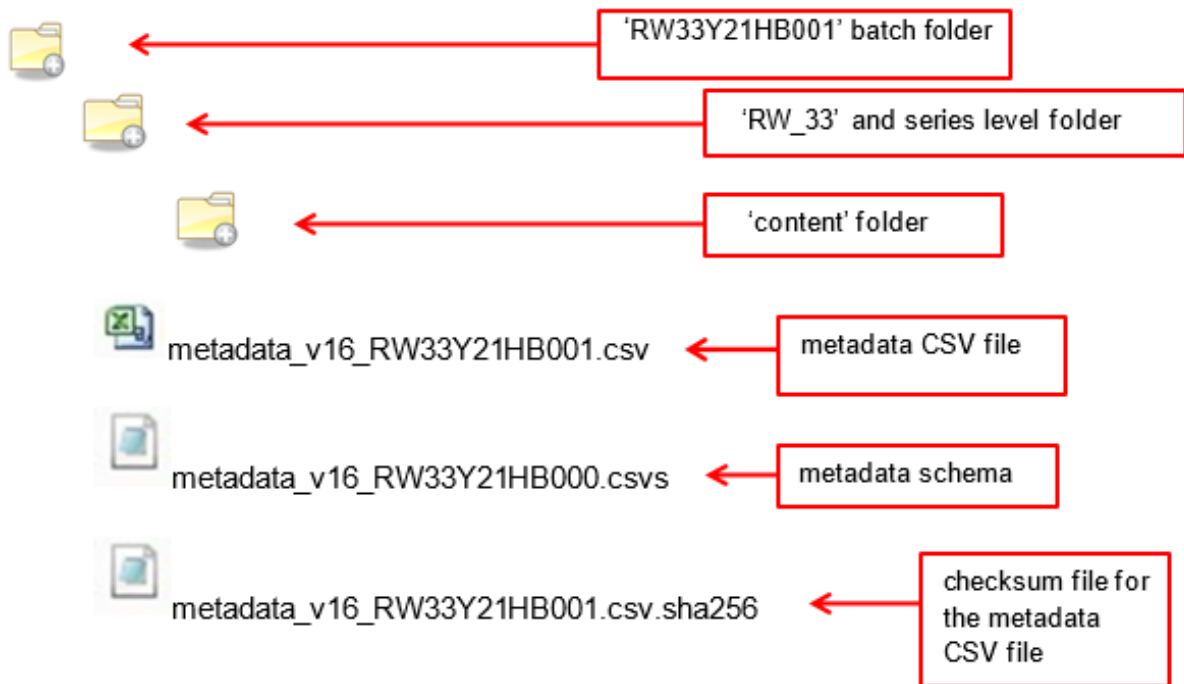


This will open up a standard 'File open' dialogue, allowing you to navigate to and select the relevant files in the file system. In the example below, 'metadata_v16_RW33Y18HB001.csv' is the 'CSV file' and 'metadata_v16_RW33Y18HB001.csvs' is the 'CSV Schema file'.



See **Appendix B** for some additional instructions on using the CSV validator and for troubleshooting.

4. Once the CSV validator has run it will either show the word PASS in the white text box or it will report errors. Copy any errors into a text editor, save and send to your Digital Transfer Adviser who can advise you how to resolve these errors;
5. Once your CSV validator reports a 'PASS' you have completed validation. Be sure to take a screenshot of your PASS screen so that you can email it to your Digital Transfer Adviser;
6. Finally, you need to generate a checksum for each metadata CSV file you have completed. Your Digital Transfer Adviser will have supplied you with an application called hash256gen.exe (either via SFTP or on the encrypted HDD). Ensure this is located within the same folder as the metadata CSV. Double click on this application. It will generate a SHA-256 file for the metadata, e.g. 'metadata_v12_RW33Y18HB001.csv.sha256'. If sending in multiple batches, you will need to run this so that you create a SHA-256 file for each metadata CSV. This allows us to confirm that no corruption occurs to the file during transfer. If you make any changes to the metadata CSV, you will need to use this application again.
7. Put these text files with their respective metadata CSV files and the schema you have used on the HDD or in your preparation area. They should sit in the department and series level folder as shown in the diagram below. Each metadata CSV file should have an accompanying SHA-256 file. (the 'hash256gen.exe' is not required).



The encrypted HDD with the validated CSV files and the digital records to be transferred is now ready to send to The National Archives. The folder structure on the HDD should replicate the structure show above.

If sending by SFTP please agree with your Digital Transfer Adviser a window for the SFTP to be open and for the transfer to be completed.

Please ensure you do not delete your records or your metadata CSV file until your records are live on Discovery and you are instructed to do so by The National Archives.

Appendix A

To add closure information you can use the table below to see how and when to populate the closure columns. If you are completing for the interim metadata CSV to send to Advisory Council you only need to complete the below closure columns before this can be submitted. Any additional metadata fields do not have to be finalised at that time.

Heading	Action you need to take to complete the template
closure_type	<p>If the file will be open on transfer, DroidConvert will have included 'open_on_transfer' in this column.</p> <p>If the file will be closed on transfer type 'closed_for' in this column.</p> <p>All folders are default open, therefore please type 'open_on_transfer' against all rows referring to folders in this column. DroidConvert will have included this for this column.</p> <p><i>Please note that the underscore between all the words must be included in order for the metadata CSV file to validate correctly.</i></p>
closure_period	<p>If the file is open, DroidConvert will have included the number '0' in this column.</p> <p>If the Advisory Council has <i>granted closure</i> for this file then type the number of years the file will be closed for, e.g. '30'.</p> <p>All folders are default open, therefore type '0' against all rows referring to folders in this column. DroidConvert will have included this for this column.</p>
closure_start_date	<p>If the file is open, DroidConvert will have left this column blank.</p> <p>If the Advisory Council has <i>granted closure</i> for this file then the start date should be the same as the date_last_modified of your record. You should copy and paste this from your date last modified column. If you know the date last modified to be invalid then they can provide a valid record date in an 'end_date' column. You should then paste this column data into closure_start_date and the closure period will be calculated from that instead of date last modified.</p> <p>All folders are default open, therefore please leave this column blank for all rows referring to folders.</p>

foi_exemption_code	<p>If the file is open, DroidConvert will have included the word 'open' in this column.</p>
	<p>If the file is <i>closed</i>, put the number of the exemption that applies e.g. '40(2)'. If multiple exemptions apply enter the exemption numbers separated by a comma but with no space e.g. '40(2),43'.</p> <p>All folders are default open, therefore please write 'open' against all rows referring to folders in this column.</p>
foi_exemption_asserted	<p>If the file is open, DroidConvert will have left this column blank.</p> <p>If the file has been <i>granted closure</i> by the Advisory Council then the date should be the date of the Advisory Council meeting where your record was granted closure. (When preparing the CSV template to send to Advisory council the date of the advisory council the application is being made to should be used.)</p> <p>The format for writing this date is 2016-12-30T00:00:00</p> <p>All folders are default open, therefore please leave this column blank for all rows referring to folders.</p>
title_public	<p>If the file is being transferred open or it's being transferred closed but doesn't have a closed file name, type the word 'TRUE' in this column. DroidConvert will have marked this column as 'TRUE'.</p> <p>If the Advisory Council has granted closure for the title of the record, type the word 'FALSE' in this column.</p> <p>All folders are default open, therefore please type 'TRUE' against all rows referring to folders in this column.</p>

title_alternate	<p>If the file is open, DroidConvert will have left this column blank.</p> <p>If the Advisory Council has approved the closure of the file name and you have typed the word 'FALSE' in the title_public column then you will need to provide an alternative file name for the record in this column. This title cannot contain any sensitivities.</p> <p>Populate the title_alternate with either 'The title of this record is closed' OR, if you want to redact part of the filename, replace the text with '[name withheld]' or '[detail withheld]', whichever is appropriate.</p> <p>All folders are default open, therefore please leave this column blank for all rows referring to folders.</p>
description_public	<p>If the file is being transferred open, or it's being transferred <i>closed but doesn't have closed descriptions</i>, type the word 'TRUE' in this column. DroidConvert will have marked this column as 'TRUE'.</p> <p>If the Advisory Council has <i>granted closure</i> for the description of the record, type the word 'FALSE' in this column.</p> <p>All folders are default open, therefore please type 'TRUE' against all rows referring to folders in this column.</p> <p><i>Please note that this field is only used if a description has been provided in addition to the file name.</i></p>

description_alternate	<p>If the file is open, DroidConvert will have left this column blank.</p> <p>If the Advisory Council has approved the <i>closure</i> of the description and you have typed the word 'FALSE' in the description_public column then an alternative description for the record must be given in this column. This description cannot contain any sensitivities.</p> <p>Populate the description_alternate with either 'The description of this record is closed' OR, if you want to redact part of the description or filename replace the text with '[name withheld]' or '[detail withheld]', whichever is appropriate.</p> <p>All folders are default open, therefore please leave this column blank for all rows referring to folders.</p> <p><i>Please note that this field is only used if a description has been provided in addition to the file name.</i></p>
-----------------------	---

Appendix B

The CSV validator also contains some additional settings which you can adjust. One of these is 'add a path substitution', this can be ignored as long as you have run the DROID export on the files in the content folder and not where the files originally sat. The file path needs to match with where the files currently sit and where the CSV validator locates them. This is why you are required to work on the files only in the content folder on your hard drive or preparation area. (Sometimes this may be required when working on a drive if the computer uses a different drive letter than when DROID was originally run, i.e 'E:/' rather than 'F:/', if this occurs your Digital Transfer Adviser can offer advice.)

Common errors to avoid when validating metadata CSV files:

- ✓ Ensure that your capitalisation is correct
- ✓ Ensure there are no extra spaces or characters before or after your entries
- ✓ Ensure the fields that should be left empty are empty
- ✓ Ensure that the fields are written exactly as per the instructions

Glossary

DROID CSV export. This provides file format information around the records, it also generates some of the information used in the metadata file.

Metadata CSV (comma separated value): is the equivalent of the paper cataloguing template for born-digital records. Each born-digital series being transferred will require its own metadata CSV file.

Batch ID: This is formed from the department and series reference, year of transfer, method of transfer (H for hard drive S and number of the batch. e.g. RW33Y21HB001 (

Schema file: The schema file is a text file which contains a set of rules which work with the CSV validator.

The code within this text document is used by the CSV validator to:

- analyse the contents of the metadata CSV file you have completed to ensure it is consistent and accurate
 - scan the original files to check their integrity
 - check that all files mentioned in the metadata are included in the folder structure in your preparation area/on your hard drive
 - check if there are any additional files in your preparation area/on your hard drive which are not represented in the metadata csv file
- Schema version number:** Transfers with different types of metadata require different schemas, the version number lets us see which version was used for the transfer. Each time a new schema is developed, it is given a new version number.

Metadata and schema filename convention: This is made up of the schema version number, series code, year of transfer, method of transfer and number of batch. For method of transfer 'H' signifies hard drive and 'F' signifies SFTP.

metadata_v16_RW33Y18HB001.csv: This would be the first transfer in 2018 sent by hard drive

metadata_v16_RW33Y18HB002.csv: Would be the second hard drive and so on.

metadata_v16_RW33Y18HB000.csvs: This is the schema filename, this always ends with B000.csvs

SFTP (Secure File Transfer Protocol), a network protocol that provides a method of file transfer. The National Archives can provide guidance on sending via SFTP.