



253 36th Street
Suite C309 #22
Brooklyn, NY 11232
<http://weareavp.com>
917.475.9630
info@weareavp.com

MDQC User Guide

Version 0.4

2018-10-10

Contact information

AVP

<http://www.weareavp.com/>

GitHub repository

<https://github.com/WeAreAVP/mdqc>

Download pages for application

Windows: <https://www.weareavp.com/wp-content/uploads/2018/10/MDQC-0.4.zip>

sha256 (.exe) :db856aedd69afa005aec86c6e3eac2be7c0754a5173931406ad15958a1436dd3

Mac: <https://www.weareavp.com/wp-content/uploads/2018/10/MDQC-0.4-os.zip>

sha256 (.dmg) :00e9d9ef271b9345b0032acc9f795d526be235b28f811727af8c9fdff2fd1dab

Special thanks:

To the wonderful folks behind ExifTool (<http://www.sno.phy.queensu.ca/~phil/exiftool/>) and MedialInfo (<http://mediaarea.net/en/MedialInfo>) for developing the great tools we built MDQC upon.

Note

This application is in beta. Please help refine it further by reporting all bugs to <https://github.com/WeAreAVP/mdqc/issues/>

Change log

Version 0.4 (October 2018)

- Fixed bug with “contain” “does not contain” filter.
- Added functionality: ability to identify specific files for QC via CSV file containing filenames/base filenames.
- Added functionality: ability to use MedialInfo files as a surrogate for media files.
- Added built-in integration with ExifTool and MedialInfo for Windows version.
- Added functionality: ability to change reports location.

Version 0.3 (May 2015)

- Fixed bug with the selected tool being saved as part of the template
- Changed behavior so that MDQC stays open after it runs
- Added “About MDQC”
- Added version number to title bar

Version 0.2 (May 2014):

- Improved dependency detection for OSX
- Improved reporting

Version 0.1 (October 2013):

- Initial release

Requirements

- MDQC is only compatible with Mac OS version 10.9 and later.
- Must use ExifTool version 10.70 or later.

Table of Contents

MDQC User Guide	1
Table of Contents	3
Brief overview and background	4
Before starting MDQC	4
For Mac Users	4
For Windows Users	4
Installing MDQC	5
Windows Users	5
Mac Users	5
MDQC interface	6
Reference file	6
Metadata Rules	7
Directory to Scan	7
Scan Rules	7
Base filenames	7
Begin Test	7
MDQC File Menu	8
Save Template	8
Load Template	8
Report Directory	8
MDQC Tools Menu	9
ExifTool	9
MedialInfo	9
Metadata Rules Window	10
Argument Pull-Down List	11
Scan Rules Window	14
Scanning MedialInfo Files	16
Base Filenames Option	18
MDQC Results	19

Brief overview and background

MDQC stands for Metadata Quality Control. It is a free open-source utility that reads the embedded metadata of a file or directory and compares it against a set of rules defined by the user. AVP created MDQC after doing a great deal of data wrangling using a combination of scripts, ExifTool, MedialInfo, spreadsheets, and formulas. The most common use to date has been to automate and minimize the time needed to QC large batches of digitized or reformatted assets. For instance, organizations have used MDQC on files being returned from a vendor or in-house lab to verify technical parameters, such as bit depth, sample rate, codec, and aspect ratio, as well as non-technical parameters such as copyright statements, title information, identifiers, and more. Although this has been the primary use of MDQC, it can be incorporated into any workflow where comparing actual metadata to expected metadata is useful. MDQC can be used on any file type supported by ExifTool and MedialInfo.

Before starting MDQC

For Mac Users

The OSX version of MDQC does not include ExifTool and MedialInfo in its install package. These will need to be installed on your system before using MDQC. The installers for these tools can be found at the following links:

ExifTool (OSX Package): <http://www.sno.phy.queensu.ca/~phil/exiftool/>

MedialInfo (i386/x86_64 CLI): http://mediaarea.net/en/MedialInfo/Download/Mac_OS

For Windows Users

The Windows install package of MDQC includes both of these tools and does not need them installed separately on your system.

Installing MDQC

Windows Users

To install MDQC on Windows, simply extract the MDQC.exe file to your desired location. You can then double-click the MDQC application file. You may have to select Run as administrator, as shown in the figure below, in order to run the application.

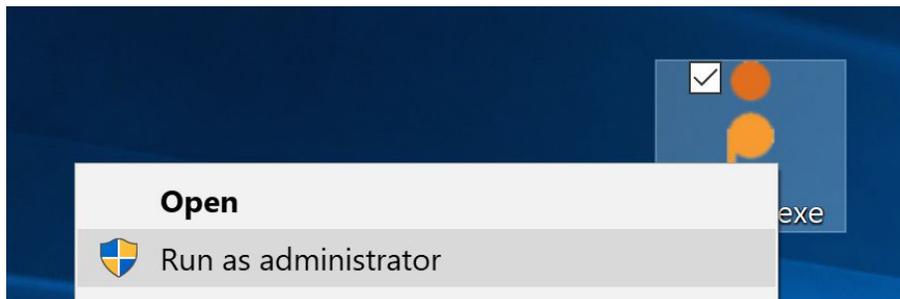


Figure 1: Running MDQC on Windows as an administrator.

Mac Users

To run the program, Mac users must extract the contents of the compressed file and simply double click on the icon to launch. It is possible that you need to allow you computer to open this application (right-click on icon, then click on “Open” to authorize app launch.)

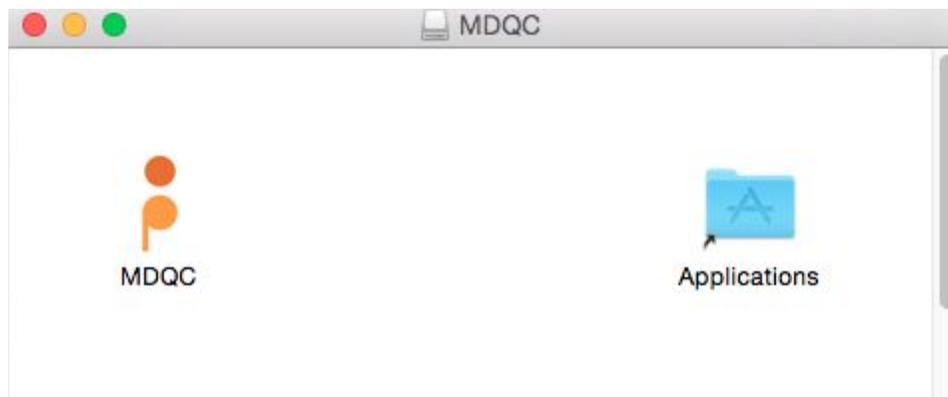


Figure 2: MDQC installation on Mac. Drag and drop MDQC into the Applications folder.

MDQC interface

MDQC has a graphical user interface (GUI). The following section details each part of the interface.

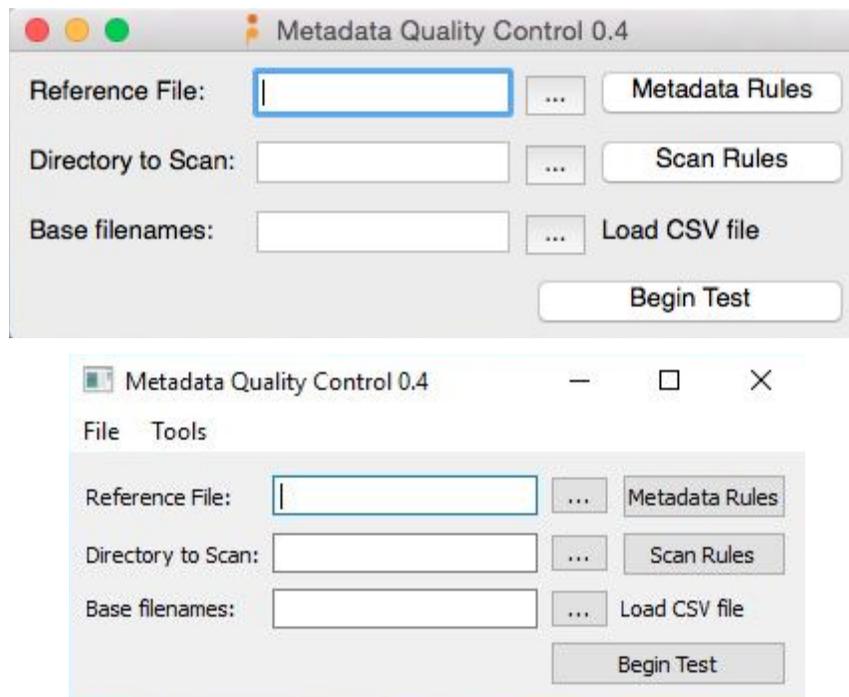


Figure 3: Screenshot of MDQC version 0.4 interface for Mac OS (top) and Windows (bottom).

Reference file

This refers to the file that you will use to set up the parameters of the metadata rules. Choose a file that: 1) Is the same type as the files you want to perform quality control on, and 2) ideally contains the metadata in it that you want. The former will ensure that the you are working with the correct structure and fields. To the extent that the reference file contains the metadata you want, the latter will populate the structure and fields with the correct values. For example, you should choose a .wav file as your reference file if you intend to scan a group of .wav files and the chosen reference .wav file should be as close to meeting your intended specification as possible in order to make it as easy as possible. If you do not have a file that contains the metadata you want to use for testing that is ok as long as it is the same type. This will just require more work when creating the metadata rules.



This button, referred to here as the “Browse” button, allows you to select a file and/or directory. Clicking this will open up a browse window where you can point MDQC to a specific file and/or directory.



NOTE

In cases when you don't have any files that have the values you want, the reference file serves the role of giving you access to the fields that exist with sample values, but allow you to update the values for a given field. In this way, the reference file provides guidance giving insights into options, structure, and syntax but allows you to put in whatever values you like.

Metadata Rules

After you have selected your reference file, click the “Metadata Rules” button. This will present you with a list of fields on the left side, the list of values found in the reference file for each of those fields on the right side, and a dropdown list of arguments that you can use for rule creation in the center. For a more in-depth look at the Metadata Rules Window, please [Metadata Rules Window](#).

Directory to Scan

Use the “Browse” button to navigate to the directory that contains the files you wish to QC. These files can be either media files or MediaInfo files (See [Scanning MediaInfo Files](#) for more details). The directory may also contain files you do not wish to QC; you can define the parameters of the specific files you wish to QC using “Scan Rules,” as described below.

Scan Rules

After you select which directory you wish to scan, you can use “Scan Rules” to identify the specific files within that directory to test. If you need to scan multiple directories, file types, or files, you can use this function to select the parameters specific to your needs. Please see the section below titled [Scan Rules Window](#) for an example scenario.

Base Filenames

There are scenarios in which users want to identify files for testing with greater specificity than “Scan Rules” allows alone. “Base Filenames” allows users to identify files by filename by uploading a CSV file containing a list of the filenames, or base filenames, to be used in the scanning and testing process. “Base Filenames” can be used alone or in combination with “Scan Rules” to provide users with a wide range of precision and complete control over the selection of files for testing. The CSV file must consist of a single column. Specification of base filenames is not required to initiate a scan. See [Base Filenames Option](#) below for an example scenario.

Begin Test

When you have defined all parameters important to your scan, click the “Begin Test” button. This will scan the directories and files you have selected and run quality control tests according to the metadata rules, scan rules, and base filenames you have defined as part of the test. After selecting “Begin Test” the results are shown in real time within MDQC in addition to a report being generated and saved as a TSV file at the end of the process.

MDQC File Menu

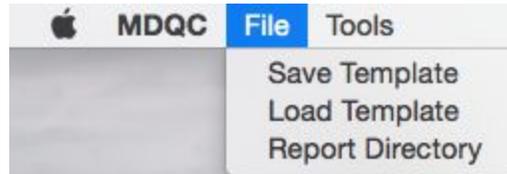


Figure 4: MDQC File Menu

Save Template

Use this option to save a template, that can be loaded at a later date, containing the metadata and scan rules that you have configured. The template does not save the filepath of the reference file or the base filenames. The reference file is no longer of use after the metadata rules are set and the base filenames, when used, will most likely be unique each time a test session is performed.

Load Template

Use this menu item to load saved templates. If the directory you intend to scan has changed since the last time you saved the template, be sure to point MDQC to the new directory using the Directory to Scan “Browse” button as described in the previous section. Also be sure to load a base filename CSV file if you wish to use this feature.

Report Directory

The “Report Directory” menu item is used to identify the directory where MDQC saves the .tsv reports. On Windows, the default setting saves the reports in the same directory as the MDQC executable file. On Mac, reports are saved by default in Documents/MDQC Reports.

MDQC Tools Menu

The user can select which underlying tool MDQC uses for setting up metadata rules and extracting/testing metadata from scanned files. The user can select from ExifTool or MedialInfo depending on what works best for the files being tested. By default, MDQC will use ExifTool.



Figure 5: MDQC Tools Menu

ExifTool

ExifTool is an open-source, free tool found at <http://www.sno.phy.queensu.ca/~phil/exiftool/>, which MDQC uses to parse the embedded metadata in files to create rule sets to use in testing. For a list of file types that ExifTool supports see <https://sno.phy.queensu.ca/~phil/exiftool/#supported>.

MedialInfo

MedialInfo is an open-source, free tool found at <http://mediaarea.net/en/MedialInfo>, which MDQC uses to parse the embedded metadata in files to create rule sets to use in testing. For a list of file types that MedialInfo supports see <https://mediaarea.net/en/MedialInfo/Support/Formats>. When using this tool, MDQC provides the user with the option to either scan the original media files or to use MedialInfo files as a surrogate for the original media files. See [Scanning MedialInfo Files](#) for more details.



NOTES

If you are a Mac user, remember to download and install these tools before running MDQC.

Metadata Rules Window

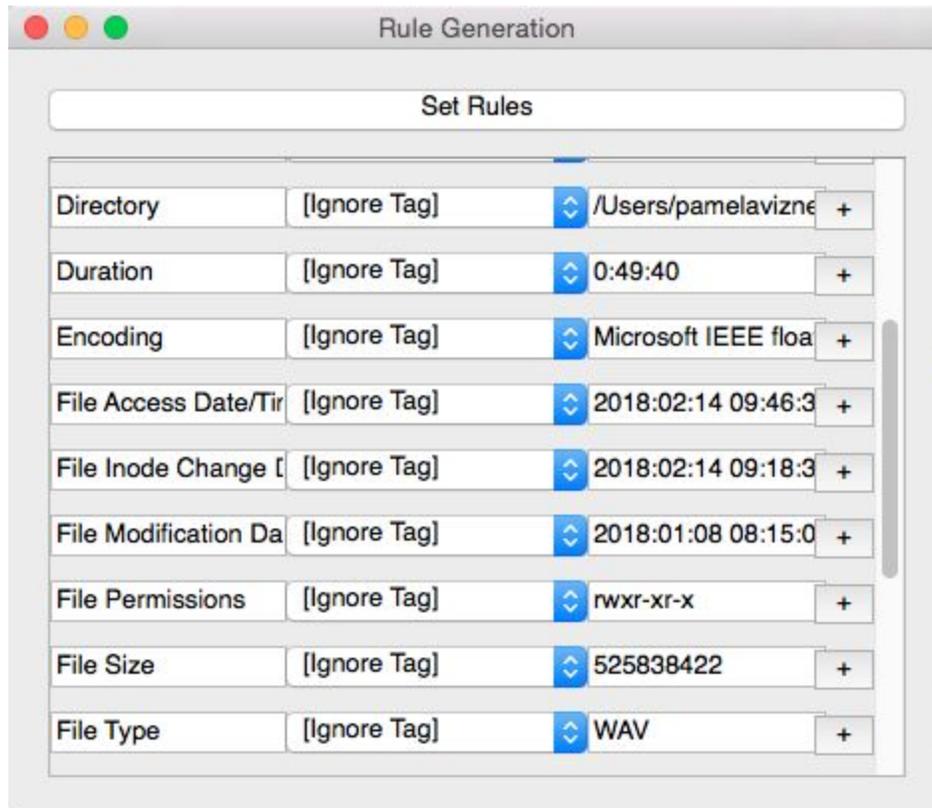


Figure 6: MDQC Metadata Rules window in Mac when a .wav file is chosen as a Reference File with ExifTool selected as the scanning method.

This is an example of the “Metadata Rules” window. Here you can choose the metadata rules you wish to use in your QC testing. The column on the left lists the fields present in your chosen reference file. We do not recommend editing the information found in this column. The center column consists of pull-down lists that allow you to select arguments for given fields (see section below for more details on arguments). The column on the right contains the fields’ values. These values, in conjunction with the arguments, will be used to create the rules that MDQC will follow in your QC tests.

Argument Pull-Down List

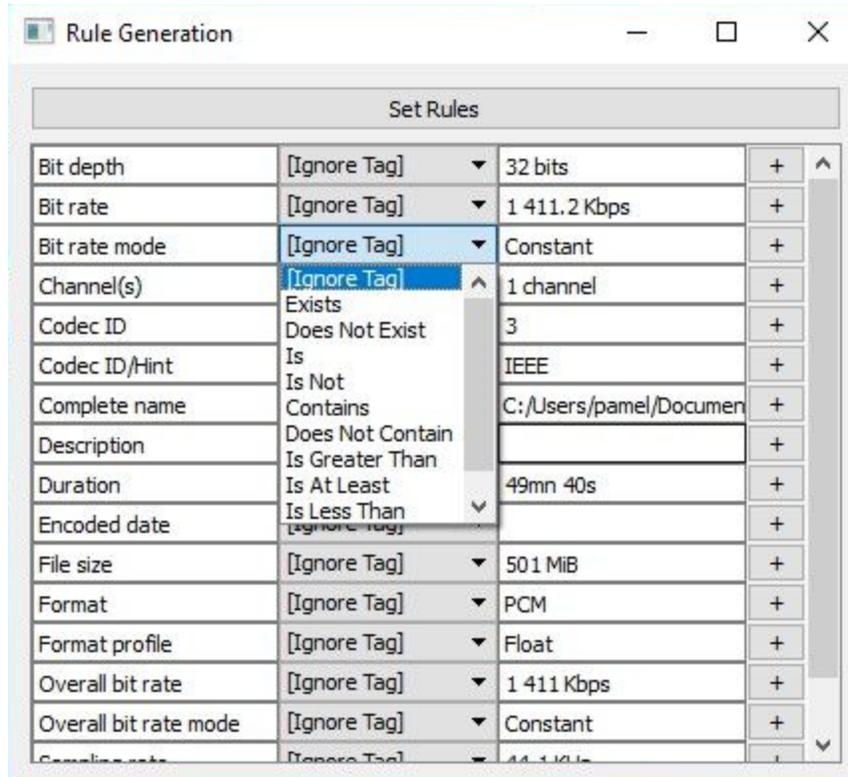


Figure 7: MDQC Metadata Rules pull-down menu in Windows when a .wav file is chosen as a Reference File with MedialInfo selected as the scanning method.

The rules in the center column are to be used as follows:

Argument	Behavior
Ignore Tag	means that a field is ignored and will not be included in the QC test.
Exists	means that a value <u>is present</u> . The specific value is not important.
Does not Exist	means that a value <u>is not present</u> . The specific value is not important.
Is	means that the value in the right-hand column of a given field <u>is exactly equal</u> to that same value and field in the files you are testing.

Is Not means that the value in the right-hand column of a given field is not exactly equal to that same value and field in the files you are testing.

Contains means that the value in the right-hand column of a given field is contained in that same field in the files you are testing.

Does not Contain means that the value in the right-hand column of a given field is not contained in that same field in the files you are testing.

Is Greater Than means that the value in the files you are testing is greater than the number in the right-hand column of a given field. MDQC will attempt to extract numbers out of values that contain numbers and letters. MDQC will allow users to input any value into fields using this argument but it will not correctly if this is not a numerical value.

Is At Least means that the value in the files you are testing is equal to or greater than the number in the right-hand column of a given field. MDQC will attempt to extract numbers out of values that contain numbers and letters. MDQC will allow users to input any value into fields using this argument but it will not correctly if this is not a numerical value.

Is Less Than means that the value in the files you are testing is less than the number in the right-hand column of a given field. MDQC will attempt to extract numbers out of values that contain numbers and letters. MDQC will allow users to input any value into fields using this argument but it will not correctly if this is not a numerical value.

Is At Most means that the value in the files you are testing is equal to or less than the number in the right-hand column of a given field. MDQC will attempt to extract numbers out of values that contain numbers and letters.



Keep in mind that some arguments might not be appropriate for all fields as MDQC will extract numbers from strings. Use your judgment when choosing this option.



The button to the right of the fields allow you to duplicate a given field in order to create multiple rules for a field. For example, if you wish to scan for a range of sample rates, you may add an additional sample rate field and use the “Is Greater Than” and “Is Less Than” arguments to establish your range.



Opening the Metadata Rules Window once "Set Rules" has been selected will only display fields and values that use an argument other than the "Ignore Tag" argument. There's currently no option to reset metadata rules once they have been set. If users want to see again all tags available for a defined reference file, they can relaunch the application or marked all fields as "Ignore Tag".

Scan Rules Window



Figure 8: MDQC Scan Rules pull-down menu in Mac.

You may choose to scan a drive or directory that contains subdirectories and/or a large number of files and file types. The “Scan Rules” window provides the following ways to test only the files you want to test:

Argument	Behavior
Directory Begins With	Indicate beginning of the name of the directory(ies) you wish to scan. This is case sensitive. Characters, numbers, and partial words are all acceptable.
Filename Begins With	Indicate beginning of the name of the file(s) you wish to scan. This is case sensitive. Characters, numbers, and partial words are all acceptable.
Directory Contains	Indicate the string that is contained in the directory(ies) you wish to scan. This is case sensitive. The character or number string must exist somewhere in the file path.

Filename Contains	Indicate the string that is contained in the file(s) you wish to scan. This is case sensitive. The character or number string must exist somewhere in the file path.
Directory Ends With	This is case sensitive. Use a character or number string that is found at the end of your directory.
Filename Ends With	This is case sensitive. Use a character or number string that is found at the end of your filename. In particular this is useful for file extensions.



This button adds additional rules as needed.

Scanning MediaInfo Files

When using MediaInfo as the underlying tool for MDQC (see [MDQC Tools Menu](#) to learn how to select tools), users can use either an actual media file (such as a .wav file) or a MediaInfo file generated from a media file as a surrogate.

MediaInfo files can be .txt or .xml and can be generated and exported using MediaInfo as a standalone application independent from MDQC.

Users can then scan collections of MediaInfo files instead of actual media files. This option can be very useful when trying to QC a large number of files that are not located on the computer running the application. It is useful for circumnavigating security and/or bandwidth issues for media files.

When MediaInfo has been selected as the scanning tool, a new dropdown titled “File Type” will pop up at the top of the window. When using MediaInfo files in place of media files - either as the reference file or as media surrogates - the “File Type” dropdown menu should be changed to “MediaInfo File”.

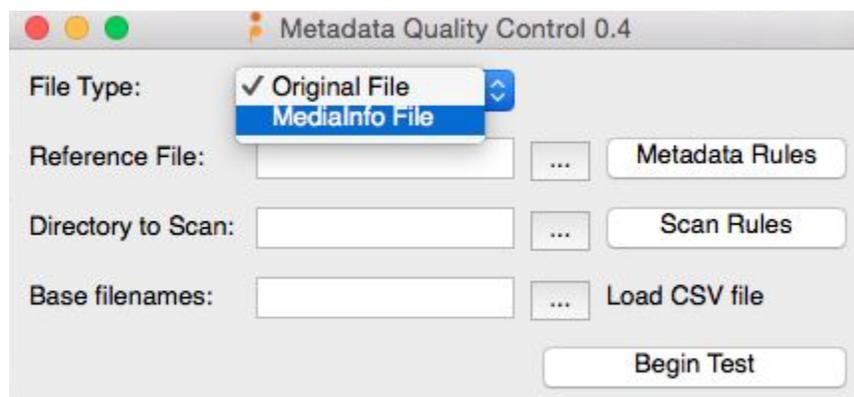


Figure 9: The File Type dropdown menu to select reference file type.

Selecting “MedialInfo file” as the file type allows the user to use MedialInfo files in place of media files for the Reference File and/or the files to be scanned. The following table illustrates all possible combinations for this feature:

Dropdown Selection	Reference File	Files to be scanned
MedialInfo File	MedialInfo	MedialInfo
MedialInfo File	MedialInfo	Media files
Original File	Media file	Media file

Figure 10: The File Type dropdown menu to select reference file type.



It is currently not possible to use a Media File as a reference if you are planning to use MedialInfo Files as surrogates. For this purpose, you can generate a MedialInfo File of the Media file you wish to use as a reference.

It is only possible to run scans using all XML or all .txt MedialInfo files. A combination of them will not return accurate results.

Base Filenames Option

The user can be more explicit about which files to scan and test than the “Scan Rules” allow through utilization of the “Base Filename” option. This allows the user to upload a list of filenames/base filenames via a single-column CSV. MDQC will then seek out all files that have a filename that “contains” any of the values specified in the CSV. This is useful when “Scan Rules” is too general to be useful and greater specificity is required. Without this feature, it would require to either segregate the files that the user wanted to test first or perform multiple passes in order to test only the files the user wanted to test.

In Figure 11, you can see a usage example of this function.

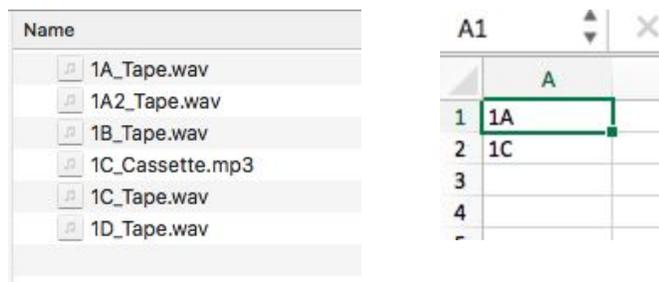
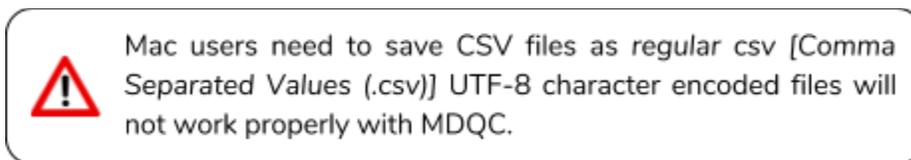


Figure 11: Usage Example 1. Files in folder to be scanned (left), .csv file containing filenames to be scanned (right). MDQC will only scan four files: 1A_Tape.wav, 1A2_Tape.wav, 1C_Tape.wav, and 1C_Cassette.mp3.



In Figure 12 below, you can see another usage example of this function in conjunction with the Scan Rules.

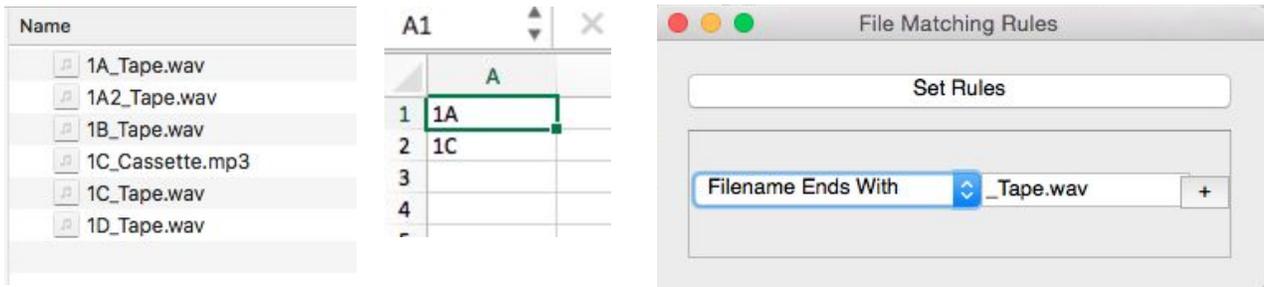


Figure 12: Usage Example 2. Files in folder to be scanned (left), .csv file containing filenames to be scanned (middle), scan rules window set (right). In this case MDQC will only scan three files: 1A_Tape.wav, 1A2_Tape.wav, and 1C_Tape.wav. Note that although 1C_Cassette.mp3 meets the rules of the csv, it does not comply with the Scan Rules set for this example, therefore it will be ignored.

MDQC Results

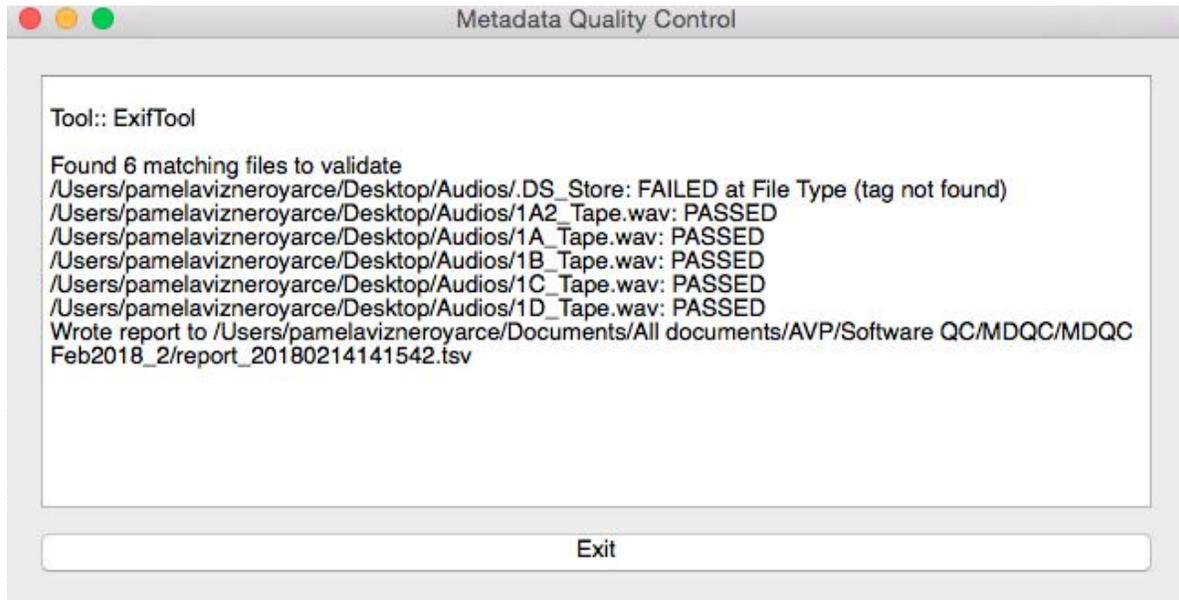


Figure 12: Results output.

Results for a test are shown in real time within the application’s window. In addition, a report is generated in the location specified within the “Report Directory” menu item. If no directory is selected, the report is saved to the same folder as the MDQC executable file. The report generated is a tab delimited TSV file.



NOTES

Rather than opening the .tsv file directly, the report should be imported into a spreadsheet application and identified as “tab delimited.” This avoids spreadsheet applications from corrupting the data, such as changing date formats, removing leading 0s, etc. Further general information about importing data into spreadsheet applications can be found here:

https://www.weareavp.com/wp-content/uploads/2017/08/Spreadsheet_Workshop_Materials-1.zip