

MDQC

MDQC User Guide Version 0.1 2013-10-27

Contact information:

AudioVisual Preservation Solutions
www.avpreserve.com

Github repository:

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

Download pages for application:

Windows:

<http://www.avpreserve.com/wp-content/uploads/2013/10/mdqc-win.zip>

Mac:

<http://www.avpreserve.com/wp-content/uploads/2013/10/mdqc-osx.zip>

Special thanks:

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

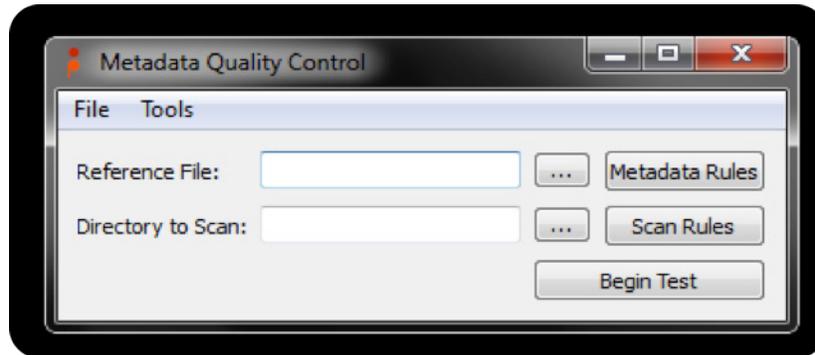
Note:

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

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. AVPreserve created MDQC after doing a great deal of wrangling data using a combination of scripts, ExifTool, MediaInfo, 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 so far MDQC can be used in any workflow where comparing actual metadata to expected metadata is useful. MDQC can be used on any file type supported by ExifTool and MediaInfo.

MDQC interface



Reference File: This refers to the file that you will use to set up the parameters of the metadata rules. Choose a file that contains all of the metadata fields you wish to QC and is representative of the file type you will be scanning. For example, you should choose a .wav file as your reference file if you intend to scan a group of .wav files. This .wav file should be as close to meeting your intended specification as possible.

 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.

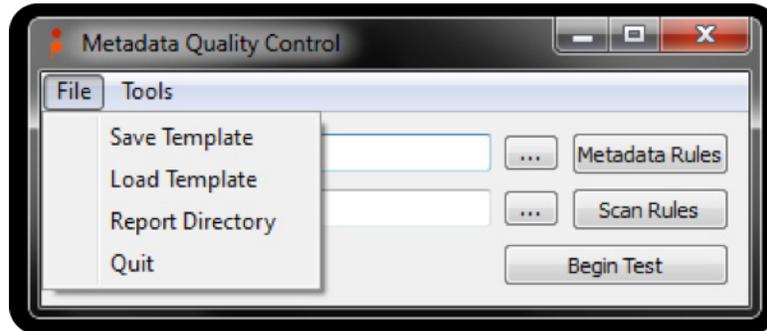
Metadata Rules: After you have selected your *Reference File*, click the *Metadata Rules* button to choose which embedded metadata fields and values you wish to use in your quality control testing. This could include specific metadata fields like file names, file types, sample rates, bit depth, or any other fields that you deem important to your QC process. For the metadata fields you don't wish to scan, simply leave the "ignore" selector as-is. For a more in-depth look at the Metadata Rules Window, please see below.

Directory to Scan: Use the *Browse* button to navigate to the directory that contains the files you wish to QC. 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 which 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* 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 you have established. The results are shown in real-time within MDQC and a report is also saved as a CSV file.

MDQC File Menu

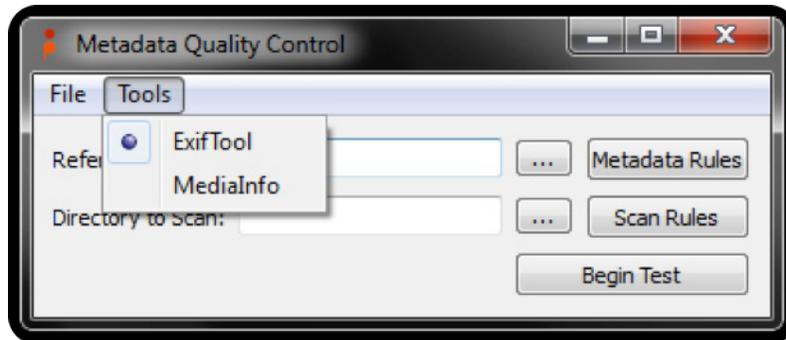


Save Template: Once you set up all relevant Metadata Rules for your intended scan, click here to save your settings for future use.

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 described above.

Report Directory: The *Report Directory* menu item is used to identify the directory where MDQC saves the CSV reports. The default setting saves the report in the same directory as MDQC.exe.

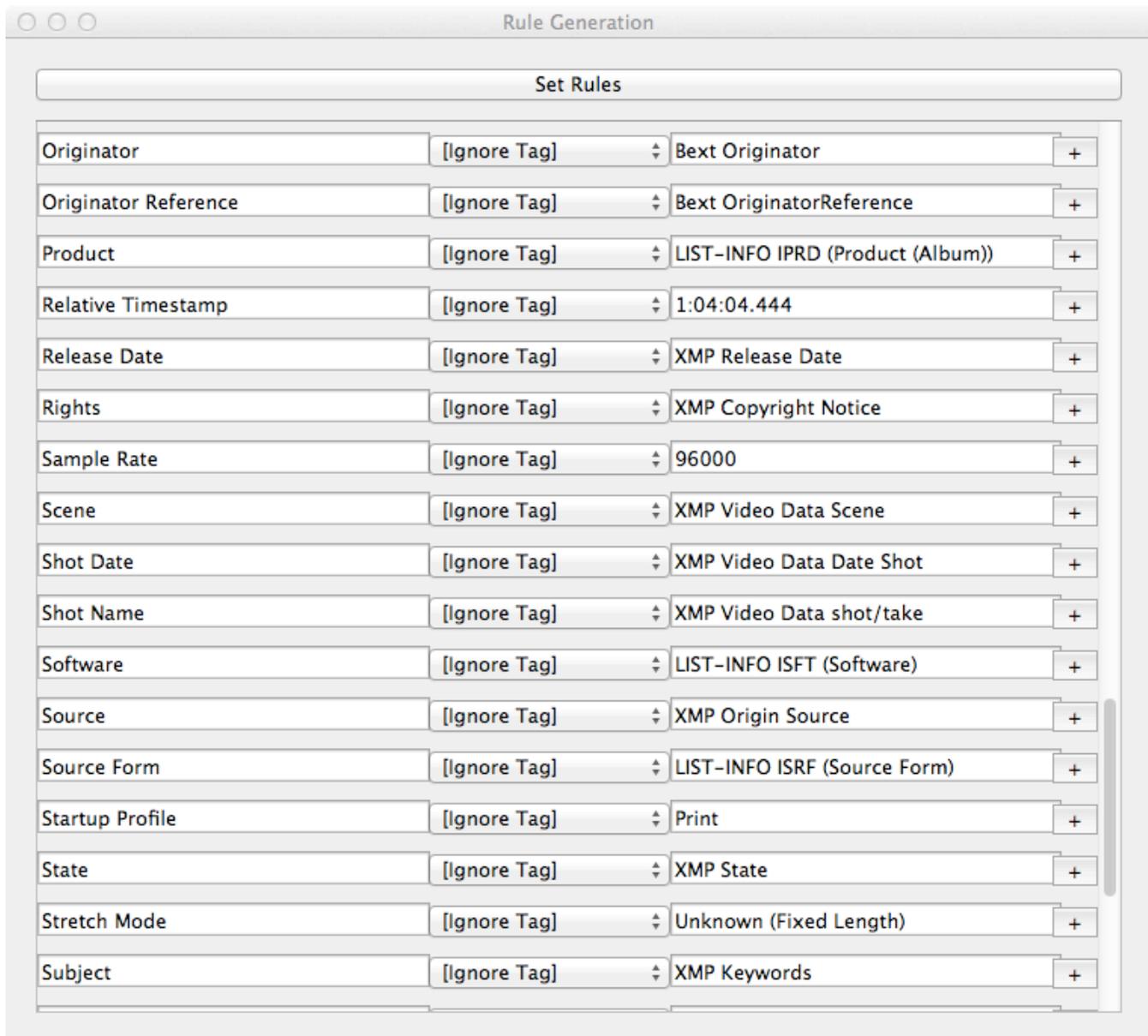
Tools Menu



ExifTool: Exiftool is an open source free tool found at <http://www.sno.phy.queensu.ca/~phil/exiftool/> that MDQC uses to parse the embedded metadata from files to create rule sets to use in testing. In Windows, make sure that the Exiftool application is located in the same directory as MDQC.exe.

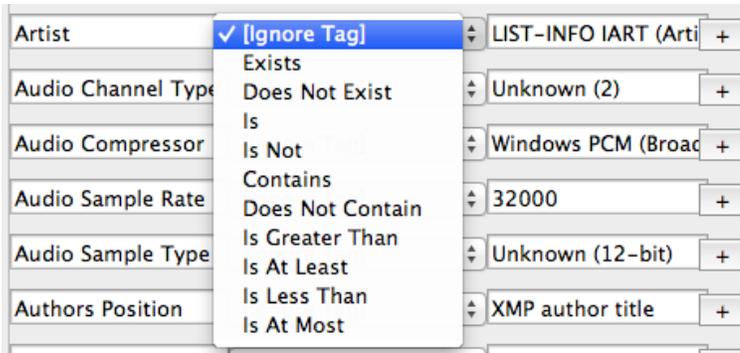
MediaInfo: MediaInfo is an open source free tool found at <http://mediaarea.net/en/MediaInfo> that MDQC uses to parse the embedded metadata from files to create rule sets to use in testing. In Windows, make sure that the MediaInfo application is located in the same directory as MDQC.exe.

Metadata Rules Window (Screenshot of the Metadata Rules window when a .wav file is chosen as the reference file.)



This is an example of the Metadata Rules window. Here you can choose the metadata rules you wish to use in your QC. 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 qualifiers for given fields. The column on the right contains the fields' values. These values, in conjunction with the qualifiers, may be used to create rules that MDQC will follow in your QC tests.

Qualifier Pull-Down List



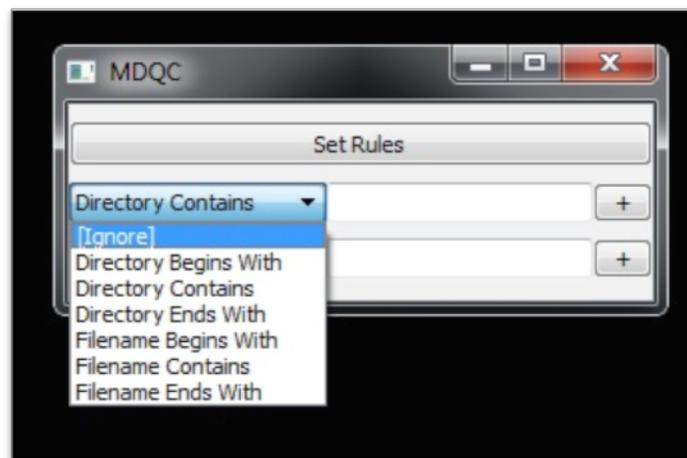
The rules in the center column are to be used as follows:

- *Ignore Tag*: means that a field is ignored and will not be included in the QC test.
- *Exists/Does not Exist*: means that a value is/is not present. The specific value is not important; if any value is present in this field, the file will pass that portion of the QC test.
- *Is/Is Not*: means that the value in the right hand column of a given field is/is not exactly equal to that same value and field in the files you are testing.
- *Contains/Does Not Contain*: means that the value in the right hand column of a given field is contained/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. This behavior may not be appropriate for all fields; use your judgment when choosing this behavior.
- *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. This behavior may not be appropriate for all fields; use your judgment when choosing this behavior.
- *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. This behavior may not be appropriate for all fields; use your judgment when choosing this behavior.

- *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. This behavior may not be appropriate for all fields; use your judgment when choosing this behavior.

The  button to the right of the fields allows 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* qualifiers to establish your range.

Scan Rules Window:

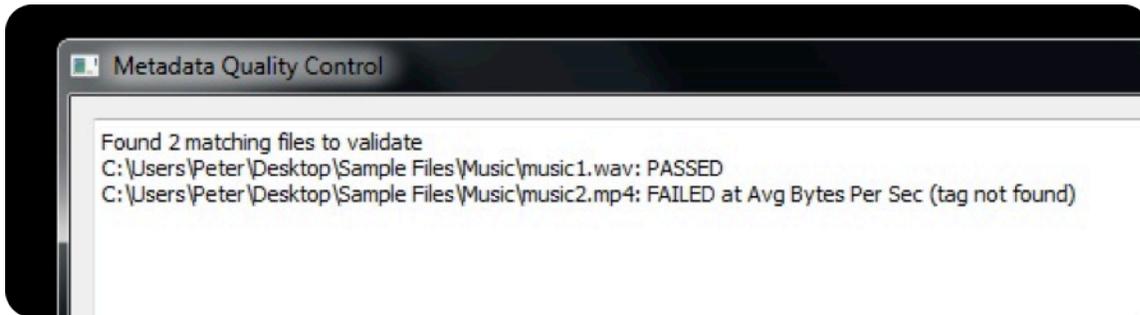


You may choose to scan a drive or directory which contains subdirectories and/or a large number of files and file types. The *Scan Rules* window provides the following ways to only test your target files:

- *Directory/Filename Begins With*: This is case-sensitive. Characters, numbers, and partial words are all acceptable.
- *Directory/Filename Contains*: This is case-sensitive. The character or number string simply must exist somewhere in the file path.
- *Directory/Filename Ends With*: This is case-sensitive. Use a character or number string that is found at the end your directory or file name (if you wish to scan only .wav files, for example, you may use this rule to do so).

The  button adds additional rules as needed.

MDQC Results:



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 MDQC.exe. The report generated is a tab delimited .csv file. Rather than opening the .csv file directly, the report should be imported into a spreadsheet application and identified as tab delimited. Further general information about importing data into spreadsheet applications can be found here:

http://www.avpreserve.com/wp-content/uploads/2013/10/Spreadsheet_Workshop_Materials.zip

Opening the file in this way will enable sorting and filtering of the columns in order to more quickly navigate the information.