

#r2b: Metadata management with ExifTool and MDQC

Amy Rudersdorf, AVP | @WeAreAVP | amy@WeAreAVP.com



**information
management
consulting &
software dev**



Free digital preservation tools

<https://www.weareavp.com/products/>

Fixity

Exactly

MDQC

Cost of Inaction Calc



Overview:
Why use these tools?

- 1. Perform basic embedded metadata review**
- 2. Automate embedded metadata review**
- 3. Ensure files conform to a set of rules**

Download

<http://bit.ly/exiftool>

ExifTool

**Download &
unzip this:**

<http://bit.ly/r2b-images>

Developed by Phil Harvey

Supports images and many AV formats

Why use it?

- Analyse embedded metadata and file system metadata --> technical and administrative metadata that may be used to recreate files in the future.
- You can view the MD or save it before you package up your files for long-term storage.

1. Open terminal window
2. Type **“exiftool”**
3. Take a quick look at help file
4. Type **“q”** to get out of help file

Let's do something with a file.

First, Let's use "Get Info" (on a Mac) to take a look at the properties of `bridge_01.jpg`. What do you see?

“More Info” section has further file information taken from the file itself

Includes:

Image information

Dimensions

Color space

Alpha channel

Now let's look at the file using ExifTool

```
$ cd /Desktop/images
```

```
$ exiftool bridge_01.jpg
```

Let's add some options from the help file to format the output a little more and make it more thorough

```
$ exiftool -G1 -D -a -e bridge_01.jpg
```

- groups by source of metadata, including system tags
- composite tags are removed

Exiftool users can export data in text, CSV, JSON, or XML formats.

```
$ exiftool -G1 -D -a -ee bridge_01.jpg > out.txt
```

```
$ open out.txt
```

You should see your text document open now and it should include the standard exiftool output as raw text.

We can do that with XML and CSV, too.

```
$ exiftool -G1 -D -a -ee -X bridge_01.jpg >  
out.xml
```

```
$ open out.xml
```


We can do that with XML and CSV, too.

```
$ exiftool -G1 -D -a -ee -csv bridge_01.jpg >  
out.csv
```

```
$ open out.csv
```

Now let's create a CSV and XML of data from all the files in the images folder.

```
$ exiftool -G1 -D -a -ee -r -csv ./ > out_all.csv
```

```
$ open out_all.csv
```

```
$ exiftool -G1 -D -a -ee -r -X ./ > out_all.xml
```

```
$ open out_all.xml
```

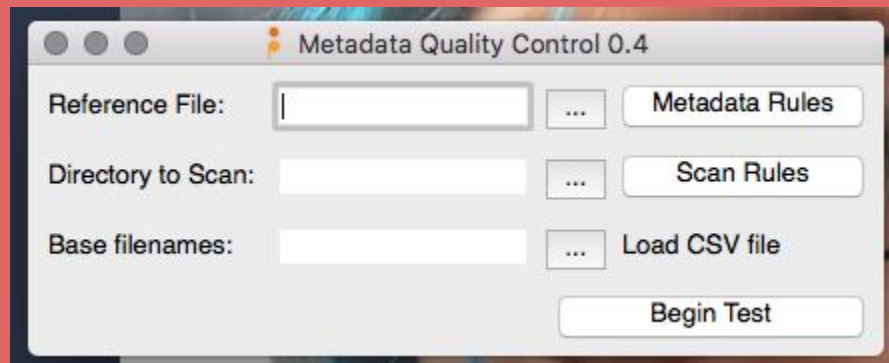
Download

[https://www.weareavp.com/
products/mdqc/](https://www.weareavp.com/products/mdqc/)

MDQC

MDQC stands for Metadata Quality Control. MDQC allows users to create rules on embedded metadata, scan a set of files, and report on the conformance of each file against the user-defined rules.

We're going to test our files to see if they conform to our rules.



1. SET UP YOUR REPOSITORY DIRECTORY

File > Report Directory

2. CHOOSE YOUR REFERENCE FILE

bridge_03.jpg

3. IDENTIFY YOUR DIRECTORY

`/whereveryouputyourunzippedfile/images`

4. SET SOME RULES

- All files should have the creator that equals “AVP”
- All files should have **any** copyright statement
- All files should be **horizontal**
- All files should be the **same dimensions**
- All files should have a **title**

!CAUTION!

**IF YOU SET THE RULES WRONG, YOU HAVE
START ALL OVER AT THE BEGINNING!!**

5. RUN THE TEST

Slide Arrange Tools Add-ons Help All changes saved in Drive

Metadata Quality Control

Tool:: ExifTool

Found 3 matching files to validate

/Users/amyrudersdorf/Desktop/images/bridge_02.jpg: PASSED

/Users/amyrudersdorf/Desktop/images/bridge_03.jpg: PASSED

/Users/amyrudersdorf/Desktop/images/bridge_01.jpg: FAILED at Creator (tag not found)

Wrote report to /Applications/report_20181014075715.tsv

Exit

METADATA QUALITY CONTROL REPORT

2018-10-14 07:58:36

METADATA RULES USED

Creator Is AVP

SCANNING RULES USED

Match all files

VALIDATION

Files found 6

Files failed 4

/Users/amyruetersdorf/Desktop/images/.DS_Store	FAILED	tag not found Creator	
/Users/amyruetersdorf/Desktop/images/#r2b_ Metadata management with ExifTool and MDQC.pptx	FAILED	tag not found Creator	
/Users/amyruetersdorf/Desktop/images/cow.jpg	FAILED	tag not found Creator	
/Users/amyruetersdorf/Desktop/images/bridge_02.jpg	PASSED		
/Users/amyruetersdorf/Desktop/images/bridge_03.jpg	PASSED		
/Users/amyruetersdorf/Desktop/images/bridge_01.jpg	FAILED	tag not found Creator	

**COWS?? WE DON'T WANT
SOME STINKIN' COWS!**

Slide Arrange Tools Add-ons Help All changes saved in Drive

Metadata Quality Control

Tool:: ExifTool

Found 3 matching files to validate

/Users/amyrudersdorf/Desktop/images/bridge_02.jpg: PASSED

/Users/amyrudersdorf/Desktop/images/bridge_03.jpg: PASSED

/Users/amyrudersdorf/Desktop/images/bridge_01.jpg: FAILED at Creator (tag not found)

Wrote report to /Applications/report_20181014075715.tsv

Exit

TRY IT YOURSELF

Free digital preservation tools

<https://www.weareavp.com/products/>



Amy Rudersdorf

amy@weareavp.com