

“Quarantine And Scrub”

A Calibre Add-On for Metadata Cleaning and Enhancement

Table of Contents

- Very Simple Example of the Q&S “Process”
- Library and Metadata Flow
- Summary of Quarantine & Scrub
- Tear-Off Menu System
- Main Menu 'ToolTips'
- Basic Steps in Using Q&S
- Scrubbing Notes
- Jobs and Job Logs
- Tag Scrubbing via Tag Rules Table
- Series Renamer/Consolidator
- Other Scrubbing Tools
- Title Word Scrubbing via Title Rules Table
- Tag Minimizer Using Priorities
- Tag, Title & Series Rules: Interlibrary Copying
- Work Tag Default by “Real” Author
- Series Word Scrubbing via Series Rules
- Work Tag Derivation From Comments
- Derive Genres Integrated With Q&S
- Work Tag Derivation from Tag Combinations
- Work Tag Capitalization Rules Table
- Work Tag String Replacement Rules Table
- Manga Awareness & Scrubbing
- Forcing Series Extraction from Title
- Deriving User Custom Columns from Title
- Dewey Decimal/Library of Congress
- Author Renaming [Pseudonyms]
- Graphic Novel/CBZ/CBR Scrubbing
- Work Tag Derivation from Book Awards Table
- Work Tag to Custom Column Derivation
- Derive a "Pseudonyms" Custom Column
- FAQs

Very Simple Example of the Basic Q&S “Process”

[Copy Real to Work](#) → [Author-Level Scrubbing](#) → [Book-Level Scrubbing](#) → [Series-Level Scrubbing](#) → [Copy Work to Real](#)

Status: **dirty**

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	S...	WorkTags	Tags
1 dirty		Washington, George Burt - A Death for a Life (Mytest Cycle, #1)	Washington, George Burt - A Death for a Life (Mytest Cycle, #1)	George Burt Washington	George Bur...			FIC006000	FIC006000

Status: **auth_ok** [if a Global/Pristine Author Found Anywhere Within Title or Elsewhere via Author Level Scrubbing]

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	WorkTags	Tags
1 auth_ok		George Burt Washington	Washington, George Burt - A Death for a ...	Washington George Burt - A...	George Burt Washington			FIC006000	FIC006000

Status: **book_ok** and Frozen [after Book Level Scrubbing regardless of any prior Status] but Real Data still dirty

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	WorkTags	Tags
1 book_ok		George Burt Washington	Washington, George Burt - A Death for a ...	A Death for a Life	George Burt Washington	Mytest Cycle [1]		Fiction:Thrillers-Espionage	FIC006000

Status: **book_ok** and Frozen and Work Data has been copied over on top of the Real Data [ready to move to your Pristine Library]

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	WorkTags	Tags
1 book_ok		George Burt Washington	George Burt Washington	A Death for a Life	A Death for a Life	Mytest Cycle [1]	Mytest Cycle [1]	Fiction:Thrillers-Espionage	Fiction:Thrillers-Espionage

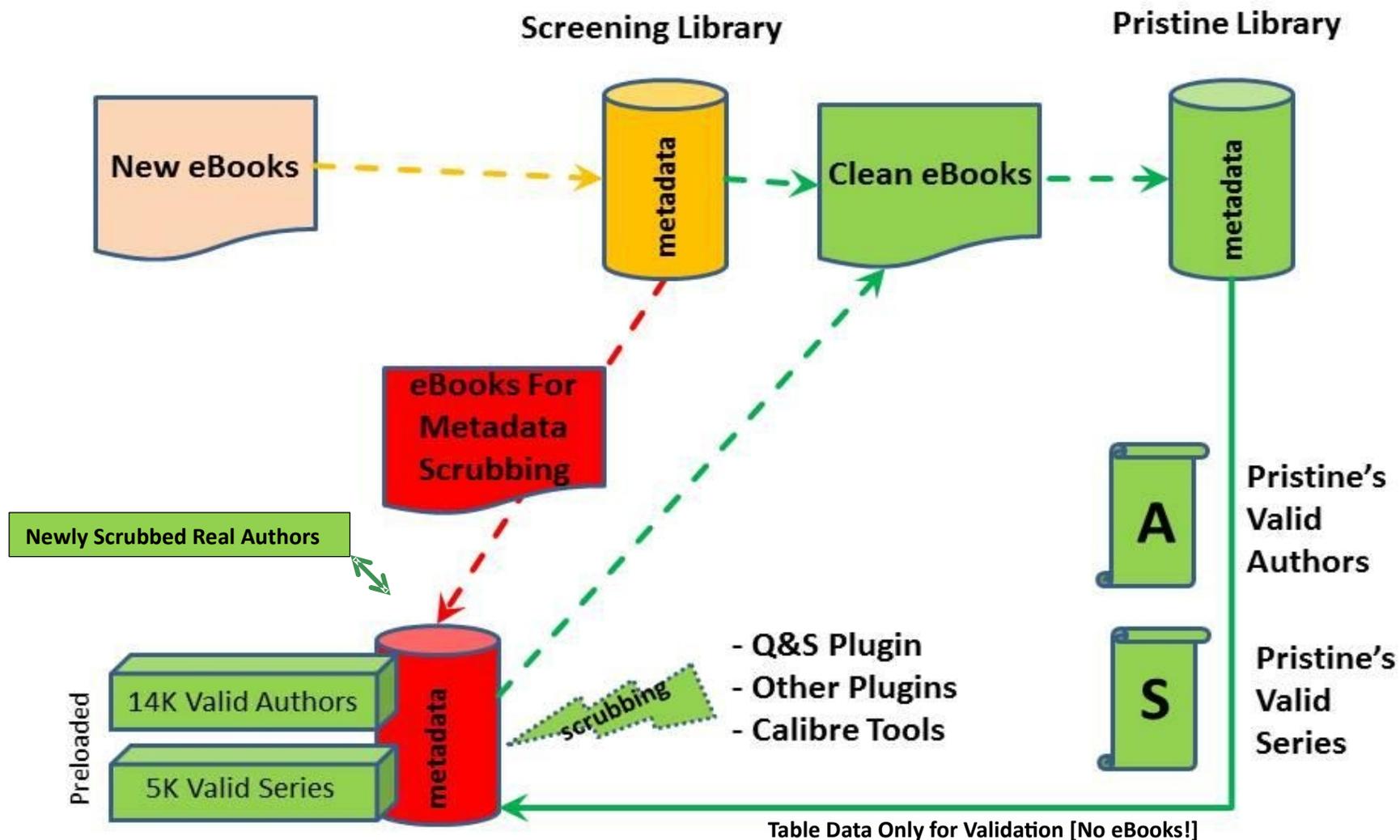
(note: this is a fake author with a fake book .txt format created with fake test metadata. 100% copyright free)

The 'status' of each book changes from '**dirty**' to '**book_ok**'. Some books will pass through an '**auth_ok**' status depending on whether or not *Author-Level Scrubbing* found anything to Scrub. After *Book-Level Scrubbing* and *Series-Level Scrubbing*, each book is ready to have the “green” Work Data copied over the “black” Real Data, totally replacing the “dirty” metadata with “scrubbed” metadata. The Work Data is then purged for those “scrubbed” books, and **only** the Standard Calibre metadata (now “**clean**”) is moved from the Special Q&S Library to the “real” Calibre Library for normal use.

A Standard Calibre Library that has **only** “**clean**” metadata is referred to as a “Pristine Library”.



Library and Metadata Flow



Quarantine And Scrub Library

Summary

Purpose: to clean malformed and/or misplaced metadata in a separate environment so you do not have to worry about your valuable “Pristine Library” being changed by this Add-on that uses many metadata.db database objects that are not known by standard Calibre.

Components: Q&S is an Add-on to Calibre that is comprised of 2 pieces: Library and Plug-In.

Library:

This custom and specifically-named Calibre Library is comprised of:

- a) One folder/directory with a name beginning with exactly “QuarantineAndScrub”.
- b) One metadata.db file.

This is a highly tailored and customized Calibre metadata.db file. It contains many SQLite Tables and Views that do not exist in Calibre. The Q&S Plugin will work only with this specific metadata.db and with this specifically named Library. The Plugin verifies that the current Calibre Library is QuarantineAndScrub before allowing the toolbar icons to become functional. It further checks for the existence of many very specific hand-tweaked custom columns and non-Calibre SQLite Tables and Views. If it does not find them, it ceases to function.

Plug-In:

The Q&S plug-in for Calibre contains all of the programming and GUI elements that interact with the special Q&S Library's metadata.db and Calibre itself.

All Menus are “Tearoff” at the Dotted Lines

Quarantine And Scrub	
PM Suggestions to Developer	
Read Q&S Instructions/User Guide	
Configure Q&S [Path to Pristine Library; Maximum for Tag Minimizer; More]	
After Version Upgrade: Create New SQLite Database Objects [Important]	
[Menu] Q&S Pristine Validation Data Maintenance	
[Menu] Q&S Purge Work Data	
[Menu] Q&S Copy Real to Work	
Scrub Work Data at Author Level [All Books]	
Scrub Work Data at Book Level [Selected Books] [Confirmation]	
Scrub Work Data at Book Level [Selected Books] [Immediate]	
[Menu] Q&S Scrub at Series Level	
[Menu] Q&S Scrub Work Tags Using Power Tools	
[Menu] Q&S Tag Rules Tables Maintenance	
[Menu] Q&S Special Tools	
[Menu] Q&S Copy Work to Real	
Derive Genres [Q&S] [Selected Books]	
Classify: DDC & LCC [Selected Books]	

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

Quarantine & Scrub User Guide

5

Purge Pristine Authors/Series Validation Tables in Q&S [If Not Really 100% Clean]	[Menu] Q&S Pristine Validation Data Maintenance
Copy Pristine Library Author/Series Tables to Q&S as Validation Data [Only If 100% Clean]	[Menu] Q&S Purge Work Data
Copy Valid Q&S Real Authors to Q&S Pristine Author Validation Table [Selected Books]	[Menu] Q&S Copy Real to Work

6

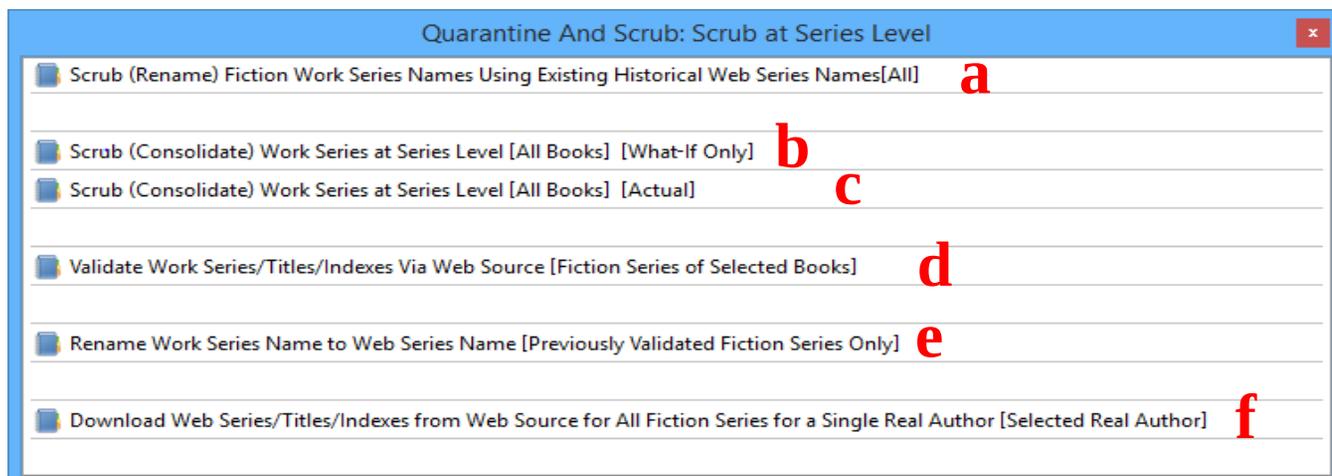
Purge All Work Data [All Books]	[Menu] Q&S Purge Work Data
Purge All Work Data [Selected Books (Max=100)]	[Menu] Q&S Copy Real to Work

7

Copy Q&S Only Dirty Tags to Work Tags [Selected Books][For Standalone Tag Scrubber]	[Menu] Q&S Copy Real to Work
Copy Q&S All Dirty Data to Work Data [All Books]	Scrub Work Title w/Author at Author Level [All Books]
Copy Q&S All Dirty Data to Work Data [Selected Books (Max=100)]	Scrub Work Data at Book Level [Selected Books] [Confirmation]
Copy Q&S Only Dirty Series to Work Series [Selected Books][Work Author & Work Title Required]	Scrub Work Data at Book Level [Selected Books] [Immediate]



Quarantine & Scrub User Guide



11

- Propagate Tags from/to Books with the Identical Series
- Change Selected Work Tags to Default Tag of 'Real' Author [Selected Books]
- Book Level Tag Scrubber [Selected Books]

12

- [Menu] Q&S Scrub Work Tags Using Power Tools
- [Menu] Q&S Tag Rules Tables Maintenance
- [Menu] Q&S Special Tools
- Q&S Scrub Using More Power Tools [All-in-One Menu]

- Auto-Populate Tags-by-Comment Table Using Tag Rules and Comments
- Tag Scrubber: 'Easy-Add' New 'Real' Tags to Purge to Tag Rules Table
- Tag Scrubber: 'Easy-Add' Work Tags-to-NEWTAG Mappings to Tag Rules Table
- Change Default Tag of 'Real' Author to Selected Work Tag [Selected Books]

13

- [Menu] Q&S Tag Rules Tables Maintenance
- [Menu] Q&S Special Tools
- Q&S Scrub Using More Power Tools [All-in-One Menu]
- [Menu] Q&S Copy Work to Real



Quarantine & Scrub User Guide

14

Quarantine And Scrub: Special Tools [Close]

-  Change Work Author (Green Only) [Single Book Only] [Spot Cleaning]
-  Change Work Title [Single Book Only] [Spot Cleaning]
-  Update Pseudonym Custom Column Using Real Authors [Selected Books]
-  Update Custom Column(s) per Table "Tags CC Mapping Control" [Selected Books]
-  Delete Non-ISBN/ISSN/OCLC Identifiers [Selected Books] **e**
-  Copy Tag & Title & Other Rules plus WSSVD From/To Previously Configured Q&S Libraries
-  Pristine Library: Reset Last-Modified Date to (Now - 24 hours) [All Books Modified in Last 24 Hours]
-  Import Q&S Reference Data from Service Pack .db

15

Quarantine And Scrub: Copy Work to Real [Close]

-  Copy WorkAuthor to Author [Selected Books]
-  Copy WorkTitle to Title [Selected Books]
-  Copy WorkSeries to Series [Selected Books]
-  Copy WorkSeries Index to Series Index [Selected Books]
-  Copy WorkTags to Tags [Selected Books]
-  Copy All Work to Real Metadata [Single Book Only]
-  Copy Valid Q&S Real Authors to Q&S Pristine Author Validation Table [Selected Books]



Main Menu 'ToolTips'

The menu items in the Q&S Main Menu are meant to be generally used in a 'top-to-bottom' sequence.

[1] Purge All Work Data [Routine Use].

[2] Copy Real Data to Work Data [Routine Use].

[3] Scrub Work Data at Author-Level.

[4] Scrub Work Data at Book-Level.

[5] Scrub Work Data at Series-Level.

[6] Perform Spot-Cleaning as Desired.

[7] Copy Work Data to Real Data.

[8] Purge All Work Data [Routine Use].

[9] Derive Genres.

[10] Derive Library Codes.

[11] Move your Cleaned Books from Q&S to your 'Real' Calibre Pristine Library.

Refer to the User Guide for details and examples.

Use Q&S in small batches of ~150 or fewer Books.

Basic Steps in Using Q&S

Step	Menu	Task
1.	3	Configure Q&S Preferences and Options
2.	4	After Version Upgrade, Create New SQLite Objects for the Q&S metadata.db
3.	5	Copy Calibre Pristine Author/Series Tables to Q&S (but only if it is truly “pristine”. Better nothing than garbage). Pristine == Clean, without defects or blemishes. Q&S is where your non-pristine books are scrubbed clean. You may not have a Pristine Library (yet). That is okay. Q&S will scrub your books so you can move them to a new Pristine Library that has only 100% clean metadata
4.	6	Purge All Work Data [All Books] [Routine Use]. Use [Routine Use] whenever possible, as it does important things that [Spot Cleaning] cannot do.
5.	Q C	Using the superb 'Quality Check' plugin, standardize Authors to be Firstname Lastname, and to have initials in the format “J.R.”. You can change them however you wish before moving the scrubbed books to your Pristine Library, but while being scrubbed in Q&S , they need to be standardized as described above. Bear in mind that the Global Authors are standardized as above, so if your Pristine Authors are different, after the Pristine Authors are copied back into Q&S the ability of Q&S to validate Authors will be diminished.
6.	7	Copy All Q&S Real Data to Work Data [All Books] [Routine Use]. The Status will be set to “ dirty ”.

Basic Steps in Using Q&S

Step	Menu	Task
7.	8	Scrub Work Data at Author Level [All Books]. The Status of changes will be set to “ auth_ok ” for only those books that were changed by Author-Level Scrubbing. The rest of the books will retain their “ dirty ” status.
8.	9 or 10	Scrub Work Data at Book Level. The Status will be set to “ book_ok ”. Although you may select a single book for Book-Level Scrubbing, all books in the Q&S Library that have a status of “ book_ok ” will automatically undergo “Miscellaneous Scrubbing” at the end of Book-Level Scrubbing. So, there is an incentive to move all “clean” books to your Pristine Library as soon as they are finished. Hence, it is strongly recommended to “Scrub in Batches”. The batch size should be about 150 or less. Less is better.
9.	11-a	Scrub (Rename) Fiction Work Series Names [All] Using Existing Historical Web Series Names. This is the same as Menu Item #11-e, but very useful at this point if you have acquired new titles within the same series that you have previously acquired and validated in Menu Item #11-d.
10.	11-b	Scrub (Consolidate) Work Series at Series Level [All Books] [What-if Only]
11.	11-c	Scrub (Consolidate) Work Series at Series Level [All Books] [Actual]
12.	12	Propagate Tags From/To Books With Identical Series
13.	14-e	Delete Non-ISBN/ISSN/OCLC Identifiers [Selected Books] (but only if you wish to do so). Caution: if you use the 'Library Codes' Plugin, use it instead! Otherwise, identifiers you may wish to keep will be deleted.



Basic Steps in Using Q&S

Step	Menu	Task
14.	12-c	Book Level Tag Scrubber [Selected Books]. Only needed if you have changed the Tag Rules since Book Level Scrubbing finished; not needed otherwise since this was already done in Book Level Scrubbing.
15.	12-d	Minimize Work Tags Per Book Using Tag Priorities [Selected Books]
16.	11-d	Validate Work Series, Work Titles, Work Index via Web Source [Fiction Series of Selected Books]. This downloads what the Web Source knows about a particular Fiction Work Series, and keeps it for use by Menu Items 11-a and 11-e . This is where the “historical” comes from. This obviates the need to re-download the same data from the Web Source repeatedly (and burdening its servers) unless you wish to update the downloaded Web Source data for newly released books in a fiction series that is already represented (<u>and previously scrubbed by Q&S</u>) in your Calibre Pristine Library.
17.	11-e	Rename Work Series Name to Web Series Name [Previously Validated Fiction Series Only]. Note that “Miscellaneous Scrubbing” automatically performed at the end of <u>any</u> Book-Level Scrubbing performs the identical function as 11-e. However, timing is everything. If you did 11-d after Book-Level Scrubbing, then 11-e could not have been done during “Miscellaneous Scrubbing”. The Q&S “Process” is recursive when it comes to acquiring, using and reusing “Web Source Series Validation Data” (WSSVD).

Step	Menu	Task
18.	12 13 14	Spot-Clean any of the Work Data you feel is warranted using any of the many specialized Q&S menu items meant for that purpose. If desired, derive Pseudonyms and/or update Custom Columns using “Tags CC Mapping Control ”.
19.		Confirm that the books now in Q&S are truly “clean” at this point. If not, repeat Step 18.
20.	15	Copy some or all Work Data to Real Data. Pick and choose. Copy all, some or none. It is up to you.
21.	6	Purge all Work Data.
22.	16	Derive Genres (if desired). DG uses Real Tags only.
23.	17	Classify: DDC & LCC (if desired). Skip this step if you use the 'Library Codes' plugin, as you should do later this directly in your Pristine Library.
24.		Using Standard Calibre: Select All Clean Books; Right Click; Copy to your 'Pristine' Library (Delete after Copy). Be sure to <u>only</u> copy Custom Columns from a Q&S Library that <u>you</u> created (e.g. Dewey Decimal Code, Library of Congress Code, OCLC-OWI, etc.). If you accidentally copy them elsewhere, delete them immediately using Calibre > Preferences. Exceptions: Genre and Pages (these are "Normal"). <i>The Q&S “Work” Custom Columns should never be copied to any library, including to another Q&S library.</i>

Scrubbing Notes

a. The “Status” column values “**dirty**”, “**auth_ok**” and “**book_ok**” are automatically set.

b. ISBN found in the Work Author, Work Title or Work Tag columns will be updated into the real ISBN field after verification that the ISBN checkdigit is valid. Its Work origin will be deleted. Note that the ISBN can just be an ISBN with no prefix such as “isbn”, and it will still be detected.

WorkTags	Tags	ISBN
	isbn9780310735649	9780310735649

WorkTags	Tags
Fiction:Action&Adventure	FIC002000
Fiction:Fantasy-Historical	FIC009030

c. Tags such as “FIC009060”, which are BISAC Subject Codes, will be converted into their meaning. Example: FIC009060 = “Fiction:Fantasy-Urban”. Do NOT delete these BISAC Subject Code Tags, since they are valuable

d. Keywords such as Thriller, Mystery, Romance, Vampire, etc. found in the Work Title will be created as new Work Tags prior to being removed from the Work Title. They then become available for Tag Scrubbing, which occurs afterwards.

e. Author Names found as a Tag will be deleted.

f. Series Names found as a Tag will be deleted.

g. Titles found as a Tag will be deleted.

Author Name as a Tag Has Been Removed:	Burt And Ernie
Series Name in a Tag Has Been Removed:	The Gray Golds
Title as a Tag Has Been Removed:	Renegade of Old Luna

h. ISBN10s will automatically be converted to ISBN13 when you run “Copy Q&S All Real Data to Work Data [All Books]”, and also when you run “Scrub Work Data at Book Level [Selected Books]”.

Scrubbing Notes

- i. Don't scrub "clean" books. Q&S cannot make them any better. Move (use > Copy (and then Delete) to your Pristine Library) "clean" books from Q&S soon after they have been completely scrubbed and are finished in Q&S. Or, at least Purge Work Data for just the "clean" books so Q&S will ignore them (and your jobs will run much faster).
- j. After your Pristine Library has had new Authors added, you should copy the Pristine Author data back to Q&S so the Author Level scrubbing can be more effective for your personalized set of Authors. The previous Pristine Authors will be purged first automatically. However, a standalone Purge Pristine Authors job will do so if you realize that your Pristine Library is not really all that pristine and you wish to purge the Pristine Authors and Series from Q&S. Highly recommended to import only "clean" data. Author-Level Scrubbing assumes that all Pristine Authors are valid, and if they are not, you will have very odd results indeed. Better no Pristine Authors than bad ones. "Bad" includes misspellings and format.
- k. Virtual Libraries used while scrubbing that have search criteria that scrubbing can change will cause odd messages and results after scrubbing jobs finish. Avoid this.
- l. Q&S is extremely English-specific in many regards. It is not intended for use with books that have titles and series not in English. However, Q&S only updates work data, not the real data, so feel free to attempt it anyway if you wish. You do not have to copy Work to Real.
- m. While scrubbing or copying jobs are running, the Calibre main GUI is locked. Please don't try to do anything in it except view the running jobs and their logs. Work Columns intentionally cannot be manually edited for numerous reasons. If you need to tweak the Work Columns, there are many menu choices for that purpose. Or, simply copy Work to Real and then manually edit the Real metadata. If desired, you can copy the "new" Real back to Work and scrub the book again. Or, if it is then "clean", move it to your Pristine Library.

Scrubbing Notes

- n. Book-Level Scrubbing will search the 'Historically Cumulative Web Series Detail' for each book being scrubbed that has no Work Series. If it finds a match, it will add the new Work Series and Work Series Index for the book, and will log a message as shown below.
- o. You must review each and every Job Log, note the warnings and errors, and take action.

[For Book]: Kisses of Death	[Missing Work Series Added per Historically Cumulative Web Data]: The Jacksonville Vampires	[Index]: 8
Valid ISBN10 Converted to ISBN13:0451232917 >>> 9780451232915		
Valid ISBN extracted and updated for book with tag of: 0451232917		
[For Book]: Ghost Towns	[Missing Work Series Added per Historically Cumulative Web Data]: The Jacksonville Vampires	[Index]: 9
[For Book]: Stone Houses	[Missing Work Series Added per Historically Cumulative Web Data]: The Jacksonville Vampires	[Index]: 1
[For Book]: Last Breaths	[Missing Work Series Added per Historically Cumulative Web Data]: The Jacksonville Vampires	[Index]: 11
[For Book]: Lords of Misrule	[Missing Work Series Added per Historically Cumulative Web Data]: The Jacksonville Vampires	[Index]: 5

Tag Scrubbing via Tag Rules Table

Table **_tag_rules** is a user-maintained scrubbing control table for Work Tags. It is used to change or delete specific Work Tags, as well as to change or delete Work Tags that match a Regular Expression. It can also be used to simply capitalize Work Tags that are often incorrectly capitalized.

To maintain it, you must install the “DB Browser for SQLite” application from <http://sqlitebrowser.org/> , which works on Windows, OSX and Linux. It is freeware, and far superior for manually editing the **_tag_rules** table than other tools such as 'SQLite Manager'. DB Browser for SQLite is designed just for this type of potentially high-volume table maintenance.

This table is used by Book Level Scrubbing as follows:

- 1) It searches the table for a simple match for each Work Tag for each book. If it finds a simple match, and a NEWTAG is specified, the Work Tag is changed to be identical to that NEWTAG. Since you do not want to delete the Work Tag, the Purge Tag column should be set to False, meaning do NOT purge the tag.

True = 1; False = 0.

	id	oldtag	newtag	purgetag
	Filter	Filter	Filter	0
1	1	/Short Stories/	Fiction:ShortStories	0
2	25	paranormal romance series	Fiction:Romance-Paranormal	0
3	39	alpha male	Fiction:Werewolves	0
4	54	Fiction - Upper Grade	Fiction:YoungAdult	0
5	55	Fiction - Upper Middle Grade	Fiction:YoungAdult	0
6	58	chocolate erotic	Fiction:Romance	0
7	99	sf	Fiction:Science	0
8	130	Magick	Fiction:Fantasy-Magic	0
9	134	Ages 8 & Up	Fiction:Children	0
10	142	erotic romance	Fiction:Romance-Erotic	0
11	149	Christian fiction	Fiction:Christian	0
12	151	/sci.*[f][iy].*/	Fiction:Science	0
13	176	/ages\s[0-9]+\s[and]*\s[&]*\s\...	Fiction:Children	0
14	177	/ages\s[0-9]+\s+[0-9]+/	Fiction:Children	0
15	178	/grades\s[0-9]+\s+[0-9]+/	Fiction:Children	0
16	179	/children.+books/	Fiction:Children	0
17	180	/large\stype\sbooks/	Fiction:Children	0
18	181	/juvenile/	Fiction:YoungAdult	0
19	183	/.*ya\s0.* /	Fiction:YoungAdult	0
20	185	/^cia\$/	CIA	0
21	186	/^fbi\$/	FBI	0
22	187	/^nsa\$/	NSA	0
23	188	/^wtc\$/	WTC	0
24	189	Supernaturals	Fiction:Fantasy-Supernatural	0
25	195	/Fiction:Thrillers-Espionage/	Fiction:Suspense&Thrillers	0
26	197	/Relato/	Ficción:Relato	0

Tag Scrubbing via Tag Rules Table

2) If a simple match is NOT found, then the Book Level scrubbing job will search through all of the Regular Expressions in the **_tag_rules** table. If it finds a match, it will do what the rule says to do (i.e., change it or purge it).

Afterwards, it will quit searching the table for that specific Work Tag. In other words, it will use the very first match from the table, and then move on to the next Work Tag for that specific book.

3) To learn Regular Expressions (a.k.a. regexes), I recommend <https://pythex.org/>. Learn online by rapid, interactive trial and error.

4) The Regular Expressions in the **_tag_rules** table MUST begin and end with a slash just like this: “/”. That is so the scrubbing job can easily and accurately distinguish between a “simple” OLDTAG and a regex OLDTAG. It does not use the first and last “/” in the OLDTAG column except to find regexes within the OLDTAG column. If no leading “/”, then the rule will be treated just like a simple OLDTAG for matching, and likely will never be used at all.

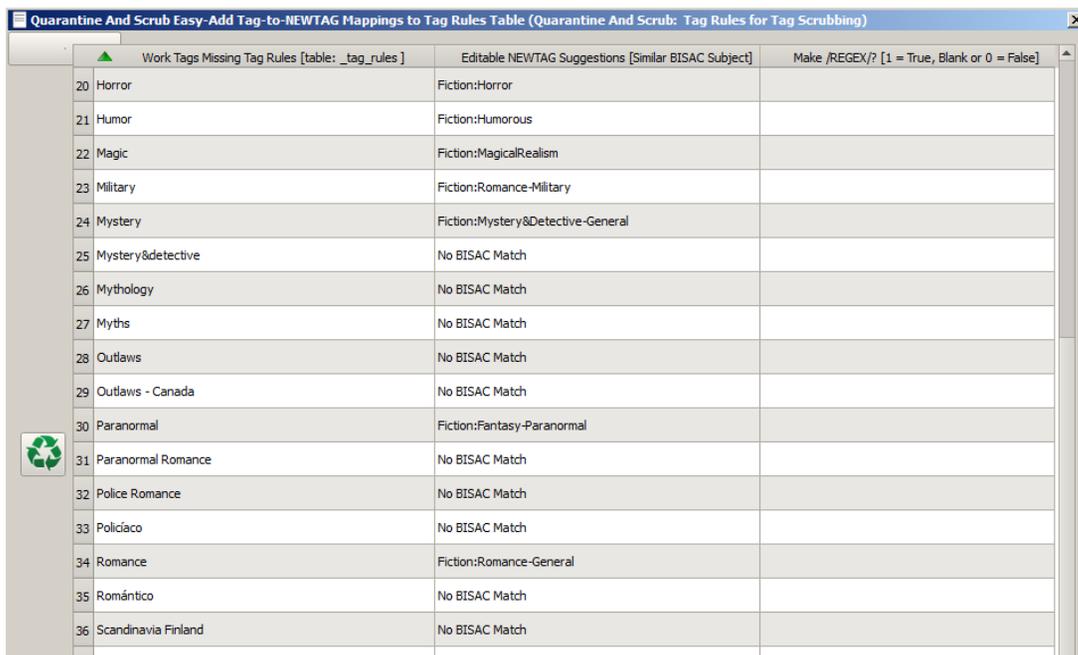
id	oldtag	newtag	purgetag
159	/ψ/		1
160	/visit\s/		1
161	/pdf/		1
162	/to-read/		1
163	/ages\s[0-9]+\s[and]*\s*\sup/	Fiction:Children	0
164	/ages\s[0-9]+-[0-9]+/	Fiction:Children	0
165	/grades\s[0-9]+-[0-9]+/	Fiction:Children	0
166	/children.+books/	Fiction:Children	0
167	/large\s\type\sbooks/	Fiction:Children	0
168	/juvenile/	Fiction:YoungAdult	0
169	/. *lang:.* /		1
170	/. *ya\s0.* /	Fiction:YoungAdult	0
171	/. *updated.* /		1
172	/^cia\$/	CIA	0
173	/^fbi\$/	FBI	0
174	/^nsa\$/	NSA	0
175	/^wtc\$/	WTC	0
176	Supernaturals	Fantasy:Superna...	0
177	/updated/		1
178	/electronic book/		1
179	/publish/		1

Tag Scrubbing via Tag Rules Table

- 5) Example: “/sci.*[f][iy].*/” is a regex of “sci.*[f][iy].*” that will find any Work Tag with the letters sci, f, i or y and change that OLDTAG to the NEWTAG of “Fiction:Science”. Capitalization is ignored, so sci-fy, sci-fi, sciENCE fiCTION, SCIFI and similar tags meaning “science fiction” all will result in a NEWTAG of Fiction:Science (unless you first change the table to make it result in another NEWTAG; it is your personal table to maintain).
- 6) To cause a Work Tag to be deleted/purged, simply set the PURGETAG value to “1”, meaning “True”. Leave the NEWTAG column empty for that rule.
- 7) If you accidentally place a value into NEWTAG but also set PURGETAG to “1”, the Book Level scrubbing program will assume that you made some type of keying error, log it, and will ignore the rule entirely. So, don’t do that.
- 8) Do not touch any other tables or views in the metadata.db database. If you do, you will have to dump all of your library ebooks to disk, delete the QuarantineAndScrub metadata.db file, reinstall Q&S from scratch (meaning the original .jar file that was really a .zip file), run the Upgrade job, and then ADD your books back into Calibre. A lot of work for no benefit, so please do not mess with anything except for table **_tag_rules** and your other tables that you own and are responsible for maintaining.
- 9) EXIT from “DB Browser for SQLite” prior to doing anything within Calibre itself. “DB Browser for SQLite” will put a database LOCK on metadata.db, and then bad and annoying things will happen.

Tag Scrubbing via Tag Rules Table

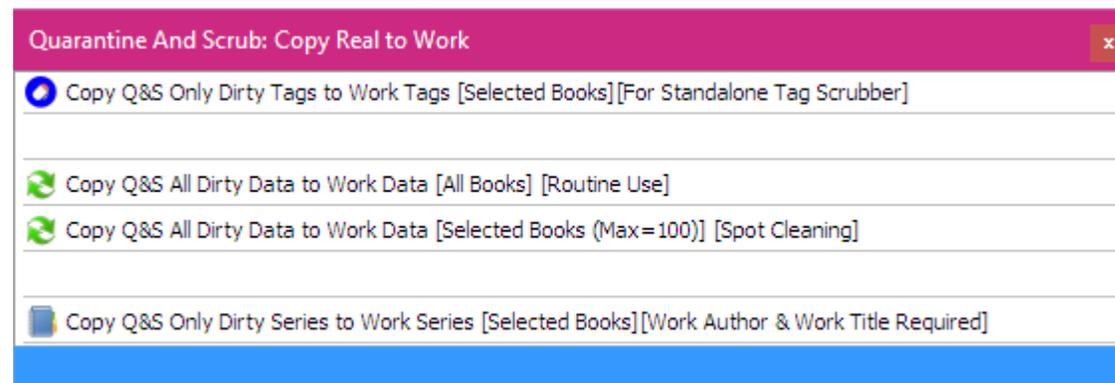
11) “Easy-Add” Tags-to-NEWTAG Mappings to Tag Rules Table. This tool does for “from old tag to new tag” change rules what the previous tool does for adding tags to purge. It finds the closest match to a BISAC Subject for each Work Tag not already in the Tag Rules Table, and displays them side by side. You can edit the NEWTAG column value to be whatever you wish (the BISAC Subject is just a hint). If you set the NEWTAG to “delete” or “DELETE”, a delete rule will be created. The 3d column controls whether a Regular Expression Tag Rule will be created, or only a Simple Match rule. Default is “Simple Match”. However, if you choose “REGEX” by using a “1” in the 3d column, Q&S will simultaneously create a “Simple Match” rule in addition to the REGEX rule to both speed processing (Simple Matches are tested first, and are much faster to process than Regular Expressions), and to make it easy for you to delete the REGEX version of the rule (without you having to manually add a new Simple Match rule to replace it) if you later realize it is has unintended consequences. Example: refer to the **Warning** on the next page.



Work Tags Missing Tag Rules [table: _tag_rules]	Editable NEWTAG Suggestions [Similar BISAC Subject]	Make /REGEX/? [1 = True, Blank or 0 = False]
20 Horror	Fiction:Horror	
21 Humor	Fiction:Humorous	
22 Magic	Fiction:MagicalRealism	
23 Military	Fiction:Romance-Military	
24 Mystery	Fiction:Mystery&Detective-General	
25 Mystery&detective	No BISAC Match	
26 Mythology	No BISAC Match	
27 Myths	No BISAC Match	
28 Outlaws	No BISAC Match	
29 Outlaws - Canada	No BISAC Match	
30 Paranormal	Fiction:Fantasy-Paranormal	
31 Paranormal Romance	No BISAC Match	
32 Police Romance	No BISAC Match	
33 Policiáco	No BISAC Match	
34 Romance	Fiction:Romance-General	
35 Romántico	No BISAC Match	
36 Scandinavia Finland	No BISAC Match	

Tag Scrubbing via Tag Rules Table

If you need to repeatedly “test and tune” your rules, you can repeatedly copy the Real Tags to Work Tags using the main menu item shown below. I highly recommend that you “test and tune” all of the entries in all Q&S tables that you personally “own”, but especially Tag Rules and Tag Priorities. The main menu item “Copy Q&S Only Dirty Tags to Work Tags [Selected Books] [For Standalone Tag Scrubber]” is intended to facilitate “testing and tuning” of your Tag Rules and Tag Priorities.

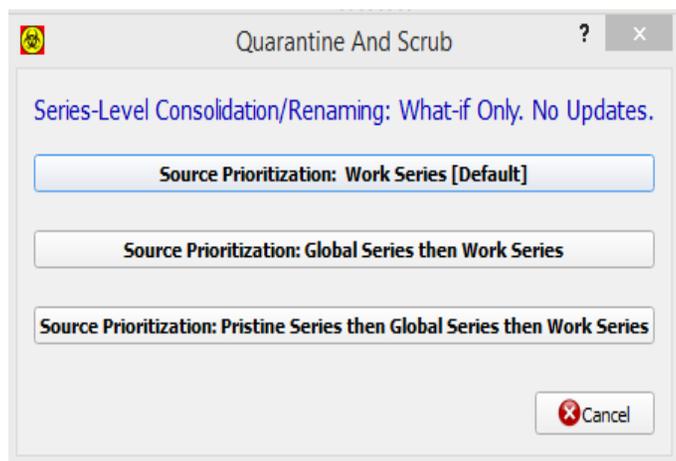


Warning: if you have a Tag Rule of “/xyz/”, then since it is a Regular Expression, any Work Tag with the letters “xyz” in it will be acted upon by your Tag Rule. Examples: Science xyz; xyz:Science; xyz:Romance; SomethingxyzHere. To narrow the rule, remove the “/”s so it will be used as a simple matching type of Tag Rule that requires an exact match. So, use “xyz” instead of “/xyz/” to avoid unintended consequences.

Example: Fiction. Avoid “/fiction/” and other extremely common words used in many different Tags, such as “Romance”, “Fantasy”, and “Science”. You should review your Tag Rules Table for problem Regular Expressions.

Series Renamer/Consolidator

This is always run AFTER Book Level has finished for ALL books in Q&S. So, all of the Work Series names have already been “Titlecased” using an English-specific algorithm that is “smart” about English grammar. Below is a multiple page example Job Log for a “What-If” (no updates) run with runtime option to only look at current Q&S book Work Series for “choices”. Runtime options: (1) Pristine, then Global, then Work Series; (2) Global and then Work Series; (3) just Work Series (a.k.a. “series”) like the example below. If your “Pristine Library” has “unscrubbed” Series Names, don’t use Option (1). This log is continued on the next page.



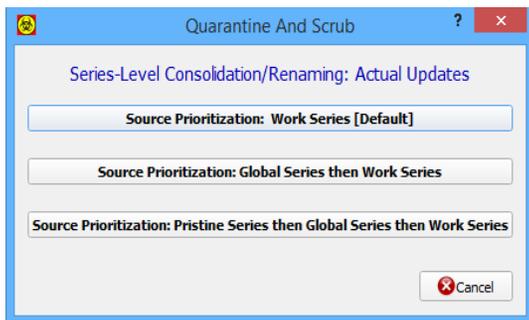
```
Starting job: quarantineandscrub
.....what-if only.....no actual updates will be made.
series source priority is: Work
S:/Users/Public/Public eBooks/QuarantineAndScrub/metadata.db
SQLite version is: 3008004
PRAGMA main.locking_mode=EXCLUSIVE;
Beginning Series Level Scrubbing
=====
For Your Reference Only
=====
Number of Global Series:      [5198]
Number of Pristine Series:   [5198]
=====
Comparing Each Work Series to Other Work Series, Global Series & Pristine Series
Single Source for Consideration is: Other Current Work Series
Selecting Appropriate Series Names
-----
Cutoff Similarity Percentage: 0.7
-----
Work Series, Other 1, Other 2, Source, Source 1, Source 2
-----
Jacksonville Vamp , Jacksonville Vampires , Jacksonville Vamp , series , series , series

Usage of Work Series as Work Series:      [1] time(s)
Usage of Other Series 1 as Work Series:    [3] time(s)
Usage of Other Series 2 as Work Series:    [1] time(s)

Similarity of Work Series to Other Series 1 is: 0.8888888888889
Similarity of Work Series to Other Series 2 is: 1.0

Action To Be Taken: Other Series 1: [ Jacksonville Vampires ] Will Replace: [ Jacksonville Vamp ]
-----
```





It may be useful to execute several times, each with a different runtime option, in the “What-If” mode until you find the best match for your particular Pristine Series and Work Series combinations. If your “Pristine Library” has books that are not really “Pristine”, perhaps you should consider moving those books to Q&S, scrubbing them using Options (2) or (3), and then moving them back to your “Pristine Library”.

Series Renamer/Consolidator

```
-----
Work Series, Other 1, Other 2, Source, Source 1, Source 2
-----
LongBlade Sea , A LongBlade Seas Novel , LongBlade Sea , series , series , series

Usage of Work Series as Work Series: [1] time(s) |
Usage of Other Series 1 as Work Series: [2] time(s)
Usage of Other Series 2 as Work Series: [1] time(s)

Similarity of Work Series to Other Series 1 is: 0.727272727273
Similarity of Work Series to Other Series 2 is: 1.0

Action To Be Taken: Other Series 1: [ A LongBlade Seas Novel ] Will Replace: [ LongBlade Sea ]

-----
Work Series, Other 1, Other 2, Source, Source 1, Source 2
-----
The Jacksonville Vampires , The Jacksonville Vampires , Jacksonville Vampires , series , series , series

Usage of Work Series as Work Series: [1] time(s)
Usage of Other Series 1 as Work Series: [1] time(s)
Usage of Other Series 2 as Work Series: [3] time(s)

Similarity of Work Series to Other Series 1 is: 1.0
Similarity of Work Series to Other Series 2 is: 0.909090909091
-----
Renaming Series Names |
Work Series: [ Jacksonville Vamp ] will be changed to: [ Jacksonville Vampires ]
Work Series: [ LongBlade Sea ] will be changed to: [ A LongBlade Seas Novel ]
Work Series: [ The Jacksonville Vampires ] will be changed to: [ Jacksonville Vampires ]

No Changes Made; What-if Only
Series-Level Scrubbing is complete.
Job complete.
```

Series Renamer/Consolidator

Reverse Swaps: Work Series Replaces 'Other Series 1' or 'Other Series 2'

```
-----  
Initial Cutoff Similarity Percentage: 0.7  
Minor Differences Cutoff Similarity Percentage: 0.75  
-----  
  
Work Series, Other 1, Other 2, Source, Source 1, Source 2  
-----  
The Testing , The Testing , Testing , series , series , series  
  
Usage of Work Series as Work Series:      1 time(s)  
Usage of Other Series 1 as Work Series:    1 time(s)  
Usage of Other Series 2 as Work Series:    1 time(s)  
  
Similarity of Work Series to Other Series 1 is: 1.0  
Similarity of Work Series to Other Series 2 is: 0.7777777777778  
  
Reverse Swap: Series Will Replace Other 2 due to (The/A/An) cutoff percentage although counts were identical.  
Action To Be Taken: Work Series : [ The Testing ] Will Replace Other Series 2: [ Testing ]  
-----
```

Normally, the Work Series is replaced by either another Work Series or a Global or PristineSeries that is also being used as a Work Series. Those “other” Work Series are named 'Other Series 1' and 'Other Series 2', and their sources may be either “series”, “pristine” or “global”. However, a “Reverse Swap” can occur if the “Other” Series (1 or 2) does NOT begin with a “The”, “An” or “A”, yet has met the “Minor Differences Cutoff Similarity Percentage” threshold, which is 75%. The “Initial Cutoff Similarity Percentage” threshold is 70%. The extra 5% allows for the indefinite and definite articles that may begin a series name. The logic actually confirms the existence of “The”, “An ” and “A ” in Work Series.

Example: if the Work Series starts with “The ”, but 'Other 2' does not although its Similarity Percentage is over 75%, then 'Other 2' is replaced by the Work Series (which is in reverse of what usually occurs).

Hence, a Reverse Swap.

Quarantine & Scrub User Guide

Downloading WSSVD By Author

“Web Source Series Validation Data”

Quarantine And Scrub: Miscellany - Download Work Series/Titles/Indexes from Web Source for All Series for a Single Author

```
Starting job: Quarantine And Scrub: Miscellany - Download Work Series/Titles/Indexes from Web Source for All Series for a Single Author
Library DB: S:/Calibre/QuarantineAndScrub/metadata.db
Tue Mar 24 08:49:52 2015
Python: Windows CPython 2.7.9
SQLite Version: 3.8.4
PRAGMA main.locking_mode=EXCLUSIVE;

Beginning Miscellaneous Scrubbing
Download Work Series/Titles/Indexes from Web Source for All Series for a Single Author [Single Author Only]
=====
Author Selected Was: John Grisham

Web Series for Author: Ford County
Web Series for Author: Theodore Boone

Number of Series Downloaded as WSSVD for Selected Author: [2]

Number of Series/Book Title Combinations Downloaded as WSSVD for Selected Author: [8]

This new WSSVD is now available to all Series Level Scrubbing functions.

John Grisham Ford County A Time to Kill 1
John Grisham Ford County Ford County 2
John Grisham Ford County Sycamore Row 3
John Grisham Theodore Boone Theodore Boone Kid Lawyer 1
John Grisham Theodore Boone Theodore Boone the Abduction 2
John Grisham Theodore Boone Theodore Boone the Accused 3
John Grisham Theodore Boone Theodore Boone the Activist 4
John Grisham Theodore Boone Theodore Boone the Fugitive 5

The Rename Work Series Name job now will have new WSSVD to use, as will Book Level Scrubbing for Work Title and Work Series Renaming.

=====
Job complete.
```

Quarantine And Scrub: Scrub at Series Level ✕

- Scrub (Rename) Fiction Work Series Names Using Existing Historical Web Series Names[All]
- Scrub (Consolidate) Work Series at Series Level [All Books] [What-If Only]
- Scrub (Consolidate) Work Series at Series Level [All Books] [Actual]
- Validate Work Series/Titles/Indexes Via Web Source [Fiction Series of Selected Books]
- Rename Work Series Name to Web Series Name [Previously Validated Fiction Series Only]
- Download Web Series/Titles/Indexes from Web Source for All Fiction Series for a Single Real Author [Selected Real Author] *

Title Word Scrubber

This is built into “Book Level Scrubbing”. Since it is automatic, there is no menu option for it, but you will likely want (and need) to perform Table Maintenance on table **_title_rules**. To maintain table **_title_rules**, you must install the “DB Browser for SQLite” application from <http://sqlitebrowser.org/> , which works on Windows, OSX and Linux. It is freeware, and far superior for manually editing the **_title_rules** table than other tools such as SQLite Manager. DB Browser for SQLite is designed just for this type of potentially high-volume table maintenance.

WorkTitle	Title
FBI USA RFID F**k WTC CIA NSA S**T	fbi usa rfid fuck wtc cia nsa shit
A III a V B ASAP C X IX VIII V	a iii a v b asap c x ix viii v
Too Many Spaces in a Newword Are Okay	too many spaces in a newword are okay
If a JFK RADAR SCUBA Not-so- UNICEF	ifa jfk radar scuba notso unicef

Q&S uses advanced English Titlecasing (not the generic Python function), which is why the “small words” of “a” and “in” are not capitalized (except at the beginning of the title). Refer to the *New York Times Manual of Style* for more information.

Q&S delivers about 62 entries in table **_title_rules** to get you started. Images of that table are shown on the following pages. However, some important comments and caveats:

1) Title Rules are applied to each Work Title as the very last thing done to Work Title in Book Level Scrubbing. Q&S has a tremendous amount of its own logic regarding Work Title as part of de-trashing and deriving Work Series and Work Series Index and even deriving Author from Work Title. So, the Title Rules are used only after all that is 100% finished. That means that you should not waste time adding “theoretical” rules to this table. Those rules might never be used by the time the Title Rules are applied by Book Level Scrubbing at the very end of Work Title Processing.

2) Be sure to test your new rules on actual Work Titles only after Book-Level Scrubbing has finished with the respective books. Regular Expressions (regexes) are powerful. Use <https://pythex.org/> to test your new regexes on actual, scrubbed Work Titles prior to updating table **_title_rules**.

Title Word Scrubber

3) You must worry about not having enough blank spaces within each rule's NEWWORD. Do not worry about having too many, as Q&S takes care of that for you. It cannot take care of having too few, however. Q&S changes all “ “ to “ “ (double to single space) and also strips the leading and trailing spaces from the Work Title. Too few spaces lead to “camelback” words, such as “TheManAteAnApple”. Q&S routinely fixes “camelback” words, but before the Title Rules are applied, not afterwards. Remember the “after 100% finished” comment above?

4) Regular expressions have certain idiosyncrasies in how they are applied for matching. Basically, a single character is not “shared” by 2 potential match targets. That means that if you want to replace the word “nsa” with “NSA” multiple times in the same Work Title, you might have a problem as shown in this example:

Your rule is: `/[]nsa[]/` for a Work Title of: *“the nsa nsa nsa blah”*. Your result will likely be: *“The NSA Nsa NSA Blah”*. That is because the space between the first “nsa nsa” will “belong” to the first “nsa”. So, the first nsa becomes NSA. The space between the first and the second nsa has been “used up”, and is not available to find the second nsa using the identical regex rule that found the first. This is referred to as “overlapping”. Don't expect it to work, at least for now. If you need to find the second nsa, you will have to create a second rule just for it that is different than the rule used to find the first nsa.

5) Note the column “Priority” in the table images shown on the following pages. You may never need to use it. The default is “100”. Its purpose is to keep rules from stepping all over each other due to the sequence in which they are applied. All of the rules in table **_title_rules** are first sorted descending by Priority. So, a rule of 200 will be applied to each Work Title before a rule of 90, and so forth. An example:

Rule #1: You want to change “15” to “Fifteen”.

Rule #2: You want to change “5” to “Five”.

The Work Title is: “The Man Received 5 to 15 Years in Prison”.

If Rule #1 were applied first, you would get “The Man Received Five to Fifteen Years in Prison”.

If Rule #2 were applied first, you would get “The Man Received Five to 1Five Years in Prison”.

So, give Rule #1 a higher numerical Priority than Rule #2 so Rule #1 will be applied first.

Title Word Scrubber

The priorities in these images do not do anything with or influence the current rules that are shown, because they all specify a single space ([]) before and after the word.

If you were to broaden the rule by removing that limitation by changing the [] to a []*, you likely would need to carefully think through what the priorities should be to avoid different title rules stepping over each other as described above in the “5 and 15” example.

	id	priority	oldword	newword	purgework
1	1	100	/[]da[]/	" CIA "	0
2	2	100	/[]wtc[]/	" WTC "	0
3	3	100	/[]nsa[]/	" NSA "	0
4	4	100	/[]fbi[]/	" FBI "	0
5	5	100	/[]rfid[]/	" RFID "	0
6	6	100	/[]usa[]/	" USA "	0
7	7	100	/[]fuck[]/	" F**k "	0
8	8	100	/[]mc[]/	" MC "	0
9	12	100	/Notsohapp/	"Not So Happ"	0
10	13	100	/i:/	"II:"	0
11	14	50	/i:/	"I:"	0
12	15	100	/iv:/	"IV:"	0
13	16	100	/ix:/	"IX:"	0
14	17	100	/[]wtf[]/	" WTF "	0
15	18	100	/[]shit[]/	" S**t "	0
16	19	30	/[][]/	" I "	0
17	20	40	/[][]/	" II "	0
18	21	50	/[][]/	" III "	0
19	22	100	/[]iv[]/	" IV "	0
20	23	50	/[]v[]/	" V "	0

Title Word Scrubber

Table:  New Record Delete Record

	id	priority	oldword	newword	purgeword
	Filter	Filter	Filter	Filter	Filter
21	24	90	/[]vi[]/	" VI "	0
22	25	95	/[]vii[]/	" VII "	0
23	26	100	/[]viii[]/	" VIII "	0
24	27	100	/[]ix[]/	" IX "	0
25	28	90	/[]x[]/	" X "	0
26	29	100	/[]*Ifan[]/	" If an "	0
27	30	100	/[]*Ifa[]/	" If a "	0
28	31	100	/[]asap[]*/	" ASAP "	0
29	32	100	/[]radar[]*/	" RADAR "	0
30	33	100	/[]nato[]*/	" NATO "	0
31	34	100	/[]unicef[]*/	" UNICEF "	0
32	35	100	/[]scuba[]*/	" SCUBA "	0
33	36	100	/[]jfk[]*/	" JFK "	0
34	37	100	/[]rsvp[]/	" RSVP "	0
35	38	100	/[]am[]/	" AM "	0
36	39	100	/[]pm[]/	" PM "	0
37	40	100	/[]bc[]/	" BC "	0
38	41	100	/[]ad[]/	" AD "	0
39	42	100	/NotSo/	"Not-So-"	0
40	43	100	/Delete Example/		1

Quarantine & Scrub User Guide

Title Word Scrubber

	id	priority	oldword	newword	purgeword
	Filter	Filter	Filter	Filter	Filter
41	44	50	/[]v\$/	" v "	0
42	45	90	/[]vi\$/	" VI "	0
43	46	100	/[]shit\$/	" S**T "	0
44	47	100	/^fbi[]/	"FBI "	0
45	48	100	/DeleteExampleByChangingToEmptyString/	" "	0
46	49	30	/[]i\$/	" I "	0
47	50	40	/[]ii\$/	" II "	0
48	51	50	/[]iii\$/	" III "	0
49	52	100	/[]iv\$/	" IV "	0
50	53	95	/[]vii\$/	" VII "	0
51	54	100	/[]ix\$/	" IX "	0
52	55	90	/[]x\$/	" X "	0
53	56	100	/[]viii\$/	" VIII "	0
54	57	100	/btw/	"BTW"	0
55	59	50	/^v[]/	"V "	0
56	60	100	/Too Many Spaces in newword Are Okay/	" Too Many Spaces in newword Are Okay "	0
57	61	100	/Espanol/	"Español"	0
58	62	90	/Volume II/	"Volume II"	0
59	63	100	/Volume III/	"Volume III"	0
60	64	100	/Volume IV/	"Volume IV"	0
61	65	100	/Volume VI/	"Volume VI"	0
62	66	80	/Volume I/	"Volume I"	0



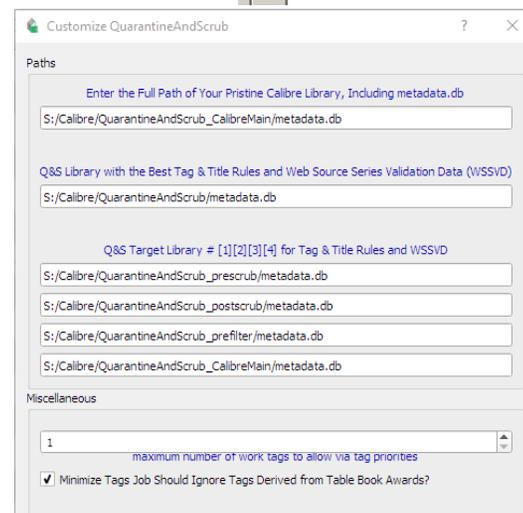
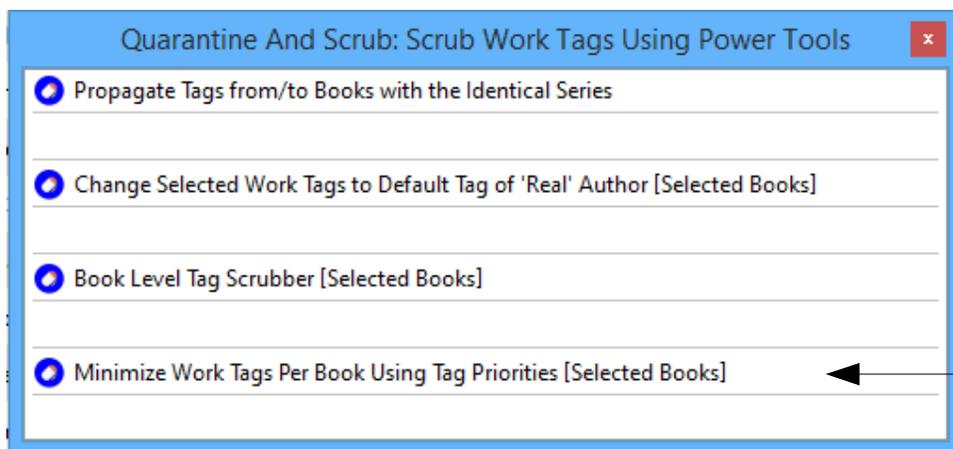
Tag Minimizer Using Priorities

This tool allows you to reduce the number of Work Tags (hopefully after you have finished with Tag Scrubbing) to a maximum that you configure in Q&S Options (just like the Pristine Library Path). The Work Tags to keep are based on their priority as defined in table **_tag_priorities**. You own that table. If a Work Tag is not in that table when this job runs, it is treated as if it has a priority of “000”. Zero. The lowest possible priority. It is logged.

[Note: seed entries are added initially by Q&S using your Work Tags (if any) plus Global BISAC Subjects to get you started.]

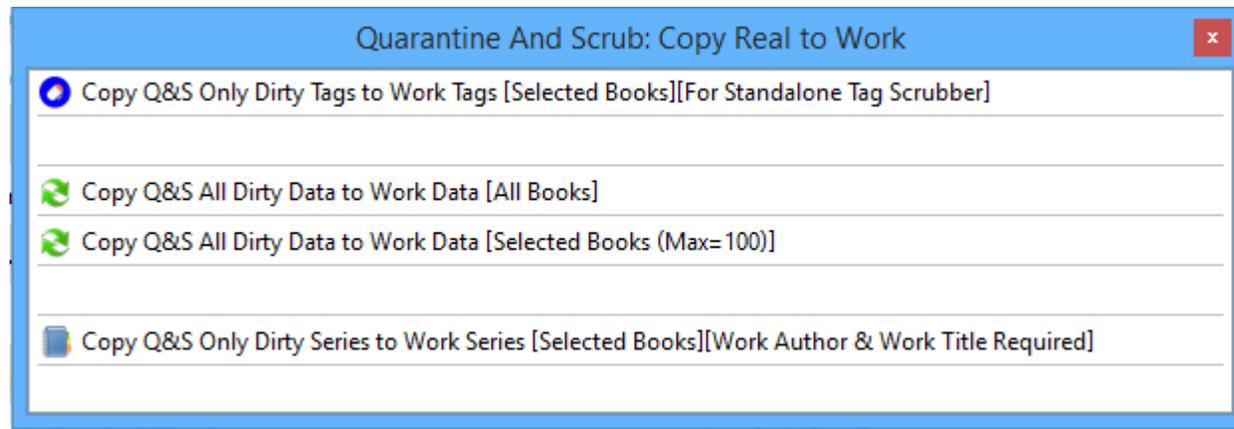
The rules are sorted “naturally” by “priority + tag” (descending), so ties of equal priority go to the tag with an alphabetic name that is “higher” in the sort. Since priorities are an integer from 0 thru 999, you can be as fine grained about them as you wish in order to avoid ties and the subsequent alphabetic tiebreakers. However, if you specify a priority greater than 999, it will be treated as if it were 999. The priorities are all made to be 3 digits (8 becomes 008) for sorting purposes. No one should need, much less want, more than 1,000 priorities.

	tag	priority
1	Test:Testing12345	600
2	testing12345	600
3	needs_fixing	500
4	Erótico	400
5	Policiaco	300
6	Fantasy & Magic	300
7	Thriller	250
8	Paranormal	200
9	Horror & Ghost Stories	150
10	Relato	115
11	testing5	105
12	testing4	104
13	testing3	103
14	testing2	102
15	testing1	101
16	Fiction:ShortStories	100
17	Fiction:Romance-Paranormal	100
18	Fiction:Werewolves	100
19	Fiction:Young_Adult	100
20	Fiction:Romance	100



Testing and Tuning Your Tag Rules

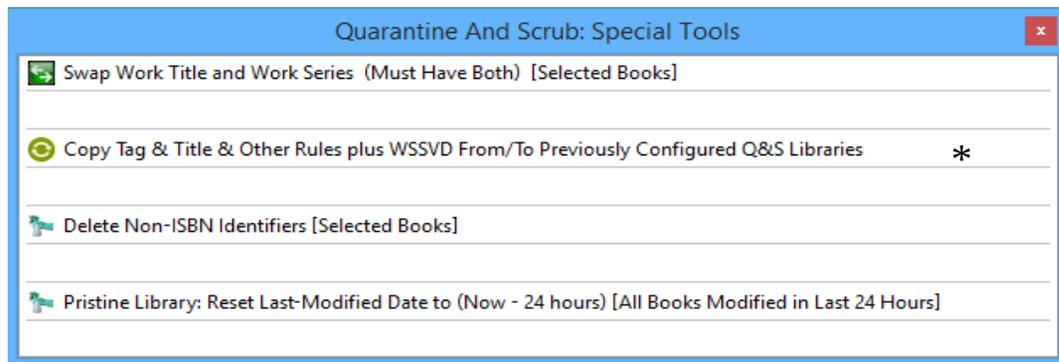
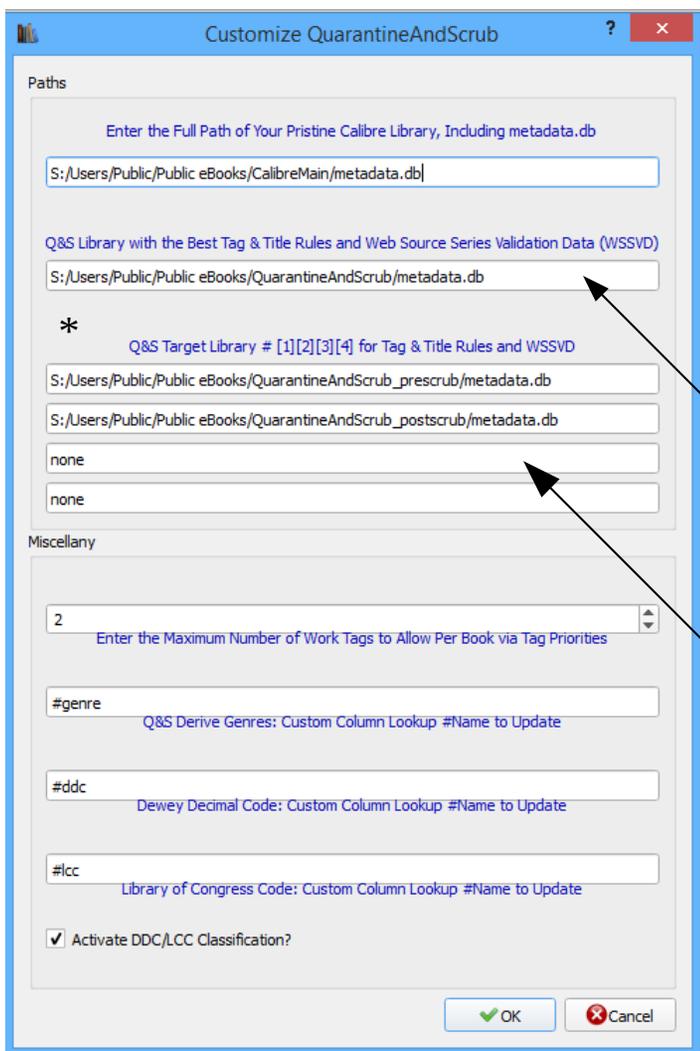
If you wish to repeatedly “test and tune” your Tag Rules (and you should), you can repeatedly copy the Real Tags to Work Tags using the main menu item shown below. It is highly recommended to “test and tune” all of the entries in all Q&S tables that you personally “own”: Tag Rules; Tag Priorities; Title Rules; and Series Rules (actually: `_tag_rules`; `_tag_priorities`; `_title_rules`; and `_series_rules`).



The next section shows how to automatically copy your “best rules” from your “best” Q&S library to your other Q&S libraries (if you have any; totally optional).

Tag & Title Rules and 'Web Source Series Validation' Historical Data

Automatic Copying Among Multiple Q&S Libraries



The Best Q&S Library path is required even if you have only one (1) Q&S Library.

If you have no Target Libraries 1-4, just use “**none**” as the path, or leave it blank. Paths 1-4 are optional.

Tag & Title Rules and 'Web Source Series Validation' Historical Data

Automatic Copying Among Multiple Q&S Libraries

```
Quarantine And Scrub
Beginning Miscellaneous Scrubbing
Copy Tag & Title Rules From/To Previously Configured Q&S Libraries

Q&S Library with the Best Tag, Title & Series Rules and Web Source Series Validation Data (WSSVD) is:  S:/Users/Public/Public eBooks/QuarantineAndScrub/metadata.db
Q&S Target Library [1] for Tag, Title & Series Rules and WSSVD is:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Target/metadata.db
Q&S Target Library [2] for Tag, Title & Series Rules and WSSVD is:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Test2/metadata.db
Q&S Target Library [3] for Tag, Title & Series Rules and WSSVD is:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Test1/metadata.db
Q&S Target Library [4] for Tag, Title & Series Rules and WSSVD is:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Target/metadata.db

Q&S Library metadata.db has been properly attached:  S:/Users/Public/Public eBooks/QuarantineAndScrub/metadata.db
Q&S Library metadata.db has been properly attached:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Target/metadata.db
Q&S Library metadata.db has been properly attached:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Test2/metadata.db
Q&S Library metadata.db has been properly attached:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Test1/metadata.db
Q&S Library metadata.db has been properly attached:  S:/Users/Public/Public eBooks/QuarantineAndScrub_Target/metadata.db

Tag Rules, Title Rules, Series Rules, Web Source Series Validation Data, Author-Tag Defaults and Pristine Authors were copied from your Best Q&S library to your other configured libraries.
=====

Number of Tag Rules Cloned from Best to Target(s):  [430]
Number of Tag Priorities Cloned from Best to Target(s):  [922]
Number of Title Rules Cloned from Best to Target(s):  [62]
Number of Series Rules Cloned from Best to Target(s):  [22]
Number of Pristine Authors Copied from Best to Target(s):  [1]
Number of Web Source Series Validation Author/Series Records Copied from Best to Target(s):  [19]
Number of Web Source Series Validation Series/Book/Index Records Copied from Best to Target(s):  [189]
Number of Author-Tag Defaults Cloned from Best to Target(s):  [479]

Cloned:  the Target table was emptied, and then all new records from the Best library were copied to the Target
Copied:  the Target table had new records added, or existing records changed to the Best library values, but no deletes.

Real Tags from all Target Libraries were copied to a Best Q&S Table used to offer choices for the 'Easy-Add' Tags-to-Purge Rules.
When you once again run this job to copy the Best tag rules back to the Targets, they will inherit any Best Tags-to-Purge Rules created using their own Tags.

Work Tags from all Target Libraries were copied to a Best Q&S Table used to offer choices for the 'Easy-Add' Work Tags-to-NEWTAG Mapping Rules.
When you once again run this job to copy the Best tag rules back to the Targets, they will inherit any Best Work Tags-to-NEWTAG Mapping Rules created using their own Work Tags.

Any Target Web Source Series Validation Series/Book/Index Records have been added to the Best Q&S Library for its own use plus future copying to all Targets.
=====
```

The Blue Stars highlight messages that convey important information that you need to be aware of.



Work Tag Default by ‘Real’ Author

Why would you want this? For books with no Work Tags, it is a very easy way to give them one.

Who maintains table **_tags_by_author_default**?

You do. Are there any “seed” entries in this table?

Yes, using your Q&S library, the most-used ‘**Real**’ Tag for each ‘**Real**’ Author not already in this table is added [*].

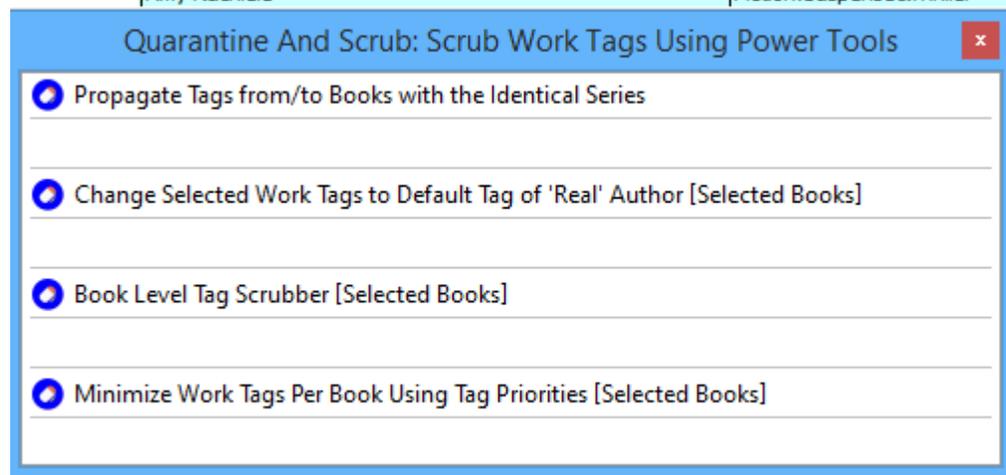
Thereafter, the Tag for each Author is not changed except by you. Q&S just inserts missing Authors so you can easily find and change them. Are these entries cloned from the ‘Best’

Q&S Library to the other ‘Target’ libraries? Yes, and the transfer is one-way only (Best to Target). How do I use this feature? Select the book for which you want the Work Tag replaced by the ‘**Real**’ Author’s Default Tag. Click the button shown above. The default Tag will totally replace whatever is in the Work Tag column for that book. [*] Anything else? Yes. Q&S runs the Work Tag through the **_tag_rules** table looking for a simple match on ‘oldtag’, and then uses the ‘newtag’ to add the missing Author-Tag combination into table **_tags_by_author_default**.

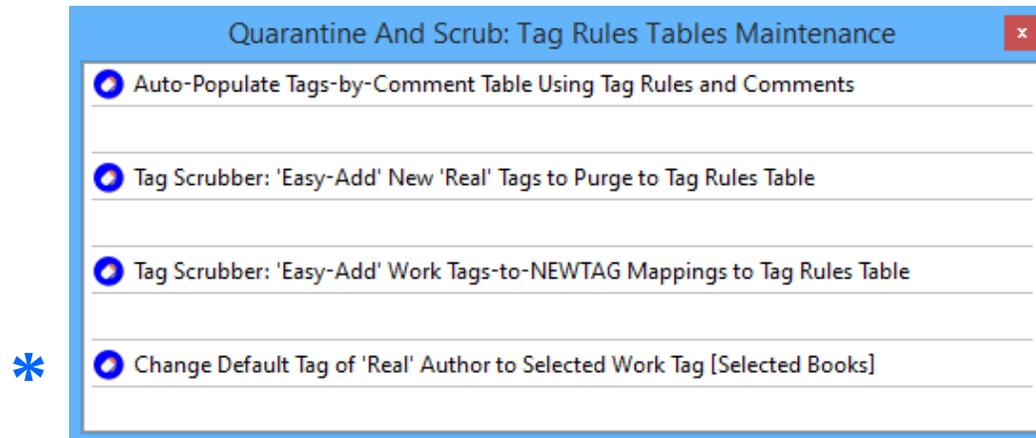
After that, it is all yours to maintain. **Remember: it uses Real Author, not Work Author.** If the Real author has not yet been overwritten with the scrubbed (**green**) Work Author, and the two are different, you should “Copy Work Author to Real Author” prior to using this feature for that particular book.

Table **_tags_by_author_default**

name	tag
Amanda J. Greene	Fiction:Suspense&Thriller
Amber Hart	Fiction:Suspense&Thriller
Amber Page	Fiction:Suspense&Thriller
Amelia Hutchins	Fiction:Suspense&Thriller
Amy Rachiele	Fiction:Suspense&Thriller



Work Tag Default by 'Real' Author



* Change Default Tag of “Real” Author to Selected Work Tag [Selected Books]

This will reset the Default Tag for the “Real” Author of any book(s) that you select prior to clicking this button. The current Work Tag (singular) will become the new Default Tag. If there are multiple Work Tags for a book that you select, then that book is entirely ignored. You may reset the Default Tag (singular) only with a single Work Tag. If you select multiple books with the identical Author but different Work Tags, the last one processed wins. This updates table **_tags_by_author_default** immediately. Remember that you always have the option to *manually* maintain table **_tags_by_author_default** instead of using this function.

Series Word Scrubbing via Series Rules Table

The instructions for this table are virtually identical to the User Instructions above for the Title Rules Table for Title Word Scrubbing. The only difference is that this table is used for Work Series instead

id	priority	oldword	newword	purgeword
1	100	/[]am[]/	AM	0
2	100	/[]pm[]/	PM	0
3	100	/[]ad[]/	AD	0
4	100	/[]bc[]/	BC	0
5	100	/[]cia[]/	CIA	0
6	100	/[]fbi[]/	FBI	0
7	100	/^fbi[]/	FBI	0
8	100	/[]wtc[]/	WTC	0
9	100	/[]nsa[]/	NSA	0
10	100	/[]usa[]/	USA	0
11	100	/[]fuck[]/	F**k	0
12	100	/[]mc[]/	MC	0
13	100	/[]shit[]/	S**T	0
14	100	/NotSo/	Not-So-	0
15	100	/Delete Example/		1
16	100	/DeleteExampleByChangingToEm...		0

of Work Title. You edit **table _series_rules** the same way that you edit table **_title_rules**, table **_tag_rules**, table **_tag_priorities** and table **_tags_by_author_default**.

The priorities etc. work the same for this table as for table **_title_rules**. Work Series Word Scrubbing via the **_series_rules** table occurs at the very end of Work Series scrubbing after it is 100% finished (just like the **_title_rules** table usage). Whatever you do to Work Series using this table is what you will see after Book Level Scrubbing has completed. The 16 rules shown above are delivered with this table to get you started, but all future maintenance must be done by you. This is your table.

Work Tag Derivation From Book Comments

Book Level Scrubbing uses table **`_tags_by_comments`** to derive new Work Tags. If you do not want to do this, then simply do not add any rows. Q&S delivers it empty. You own this table, just like **`_tag_rules`**, **`_title_rules`** and so forth. Maintain it the same way.

This is how it works: The values in the “comment” field in table **`_tags_by_comments`** are the specific keywords that you want Q&S to search each book’s Comments with. If a keyword from the table is found anywhere within a book’s Comments, then the resulting “tag” from the table is added as a new Work Tag for that book.

This table does NOT use regular expressions. The “comment” column should be simple keywords or phrases. Do **not** use punctuation. Refer to the image above for examples. However, see the next page for the Auto-Populate capability.

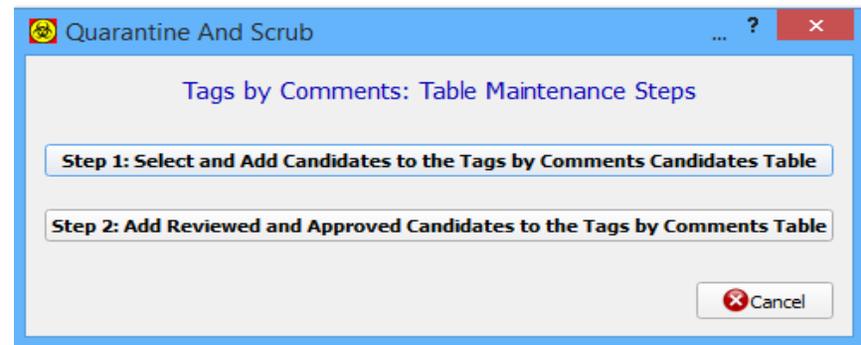
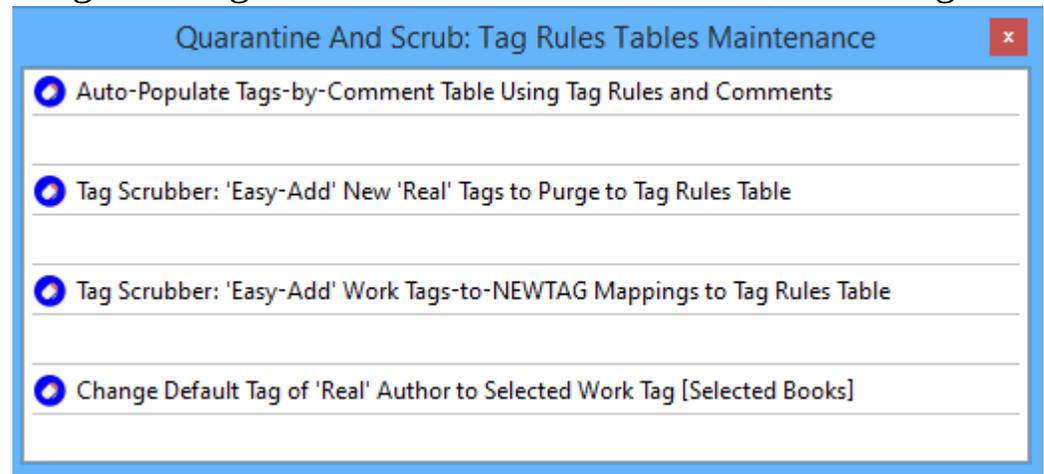
TABLE <code>_tags_by_commen</code> <input type="text"/> Search <input type="button" value="Show All"/>		
id	comment	tag
1	mystery	Fiction:Mystery
2	thriller	Fiction:Thriller
3	romance	Fiction:Romance
4	science fiction	Fiction:Science
5	werewolf	Fiction:Werewolves
6	werewolves	Fiction:Werewolves
7	vampire	Fiction:Vampires
8	shapeshifter	Fiction:Shapeshifter
9	zombie	Fiction:Zombies
10	fantasy	Fiction:Fantasy

Auto-Populate the Tags-by-Comment Table Using Tag Rules and Comments

This time-saving feature runs the Tag Rules Table “oldtags” through the *Calibre* Comments Table, searching for simple matches. When it finds a match, it uses the “oldtag” as the Comment and the “newtag” as the Tag for new entries it puts into table **_tags_by_comment_candidates**. After you review those entries, it then moves them to the real **_tags_by_comment** table. The Book Level Tag Scrubber uses that table, as does Book Level Scrubbing itself. This makes it easy to derive new Work Tags using your own Tag Rules with your own library’s book Comments.

Step 2 does not do what Step 1 does. You must execute Step 1 prior to Step 2, with a review of the candidates in between the Steps.

Work Tags as “oldtags” are ignored for this purpose if they contain the following punctuation: colon; semi-colon; period. Example: Fiction:Science as an oldtag would be ignored because it is essentially a BISAC subject, not a “real” word found in a Comment. Ditto for Factual:Neurology .

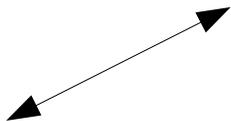


Derive Genres

The Derive Genres functionality is fully integrated into Q&S. It has its own User Guide due to its complexity. However, a brief overview of DG is provided here.

If you are interested in Deriving Genres, you can download the DG User Guide from the Q&S Original Post on www.mobileread.com.

All Q&S libraries come 'standard' with a Custom Column with a look-up name of "#genre". Ditto for "#pages".



```
Quarantine And Scrub
Starting job: Quarantine And Scrub
S:/Users/Public/Public eBooks/QuarantineAndScrub_DeriveGenres/metadata.db
Tue Jan 20 11:51:19 2015
Python: Windows CPython 2.7.9
SQLite Version: 3.8.4
SQLite DB Locking Mode: EXCLUSIVE [GUI Is Locked]

=====
Q&S: Deriving Genres for 8 Book(s)
=====

Q&S Derive Genres is controlled by 4 DG tables within Q&S that are User-Maintained:
.....      _dg_nf_language_keywords
.....      _dg_genre_tag_rules_factual
.....      _dg_genre_tag_rules_fiction
.....      _dg_genre_author_rules

Q&S DG uses Real Tags, not Work Tags, so finalize Real Tags prior to execution.

Q&S DG adds missing Tag Rule Combinations to DG table _dg_missing_tag_rules.

You should use table _dg_missing_tag_rules to help you create new Tag Rules.

You should delete rows in table _dg_missing_tag_rules that no longer are useful to you.
=====

Q&S Configured Custom Column #Name to be used is: #genre
=====

Number of Selected Books to Process:                8
Number of Rules for Fiction/Poetry Processed:      129
Number of Rules for Factual Processed:             111
Number of Rules for Authors Processed:            10278
Number of Multi-Language Factual Keywords:        50
Number of Books for Which Tags Were Gathered:      8
Number of Factual Books Identified:                0
Number of Non-Factual Books Identified:            8
Number of New Genres Derived from Rules:           5
Number of Books With No Applicable Rules:         3
Number of Missing Tag Rules in _dg_missing_tag_rules: 4
=====

Genre Derivation ended normally
```

Quarantine & Scrub User Guide

Derive Genres

Table: 

	id	new_genre	tag_keyword_1	tag_operator_1	tag_keyword_2	tag_operator_2	tag_keyword_3	tag_operator_3	tag_keyword_4	tag_operator_4	tag_keyword_5
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	1	Factual:Antiques&Collectibles	factual	AND	antique	OR	collectible	NONE	NONE	NONE	NONE
2	3	Factual:Architecture-General	factual	AND	architecture	AND	general	NONE	NONE	NONE	NONE
3	5	Factual:Architecture-Historic Preservation	factual	AND	architecture	AND	historic	NONE	NONE	NONE	NONE
4	7	Factual:Architecture-History	factual	AND	architecture	AND	history	OR	historical	OR	ancient
5	9	Factual:Art-Drawing	factual	AND	art	AND	drawing	NONE	NONE	NONE	NONE
6	11	Factual:Art-Fashion	factual	AND	art	AND	fashion	NONE	NONE	NONE	NONE

Table: 

	id	new_genre	tag_keyword_1	tag_operator_1	tag_keyword_2	tag_operator_2	tag_keyword_3	tag_operator_3	tag_keyword_4	tag_operator_4	tag_keyword_5
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
19	241	Fiction:Children	fiction:children	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
20	242	Fiction:Classics	classics	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
21	243	Fiction:Classics	fiction:classics	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
22	244	Fiction:Comics&GraphicNovels	comics	OR	graphicnovels	NONE	NONE	NONE	NONE	NONE	NONE
23	245	Fiction:Comics&GraphicNovels	fiction:comics&graphicnovels	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
24	246	Fiction:Cowboys	cowboy	OR	cowboys	NOT	westerns	NOT	western	NONE	NONE
25	247	Fiction:Dystopian&Apocalyptic	dystopian	OR	apocalyptic	OR	dystopia	OR	apocalypse	NONE	NONE
26	248	Fiction:Dystopian&Apocalyptic	fiction:dystopian&apocalyptic	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
27	249	Fiction:Espionage	espionage	NOT	war	NONE	NONE	NONE	NONE	NONE	NONE
28	250	Fiction:Espionage	fiction:espionage	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
29	251	Fiction:Fantasy	fantasy	NOT	paranormal	NOT	steampunk	NOT	urban	NONE	NONE
30	252	Fiction:Fantasy	fiction:fantasy	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
31	253	Fiction:Fantasy-Paranormal	fantasy	AND	paranormal	OR	supernatural	NOT	horror	NOT	vampire
32	254	Fiction:Fantasy-Paranormal	fiction:fantasy-paranormal	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
33	255	Fiction:Fantasy-Romance	fantasy	AND	romance	NOT	paranormal	NOT	urban	NOT	steampunk
34	256	Fiction:Fantasy-Romance	fiction:fantasy-romance	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
35	257	Fiction:Fantasy-SteamPunk	fiction:fantasy-steampunk	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
36	258	Fiction:Fantasy-SteamPunk	steampunk	NONE	NONE	NONE	NONE	NONE	NONE	NONE	NONE
37	259	Fiction:Fantasy-Urban	fantasy	AND	urban	NONE	NONE	NONE	NONE	NONE	NONE



Quarantine & Scrub User Guide

Derive Genres

Table:

	id	author	fiction_genre	factual_genre
	Filter	Filter	Filter	Filter
1	2	a. núñez alonso	Ficción:Histórico	NONE
2	3	louis lamour	Fiction:Westerns	Factual:Cowboys
3	6	a. grove day	NONE	Factual:Biography&Autobiography
4	7	a. lee martinez	Fiction:Fantasy	NONE
5	8	a. manette ansay	Fiction:Romance-Historical	NONE
6	9	a. meredith walters	Fiction:Romance	NONE
7	10	a. michael matin	Fiction:Classics	NONE
8	11	a. scott berg	NONE	Factual:Biography&Autobiography

Table:

	id	nf_keyword
	Filter	Filter
1	1	factual
2	3	noficcio
3	4	noficción
4	5	sachbüch
5	6	não-ficção
6	7	tényleges
7	8	staðreyndir
8	9	fiorasach
9	10	gerçek
10	11	dejanska
11	13	factuel
12	14	fakta
13	15	faktaj
14	16	faktik
15	17	faktiki
16	18	faktiline
17	19	faktinis
18	20	faktisks
19	21	faktual
20	22	faktuelle

Remember: This Q&S Derive Genres uses tables that you maintain to hold your Boolean Tag Rules, Author Genre Default Rules, and the Non-Fiction Keywords. You will be using the Non-Fiction Keywords to tell Derive Genres what is Factual/NonFiction. Everything else is Fiction/Poetry, and is derived as such.

For more, download the full Q&S DG User Guide available at:
<http://www.mobilerread.com/forums/showthread.php?p=2940048#post2940048>



Work Tag Derivation from Tag Combinations

rowid	tag_keyword_1	tag_keyword_2	tag_keyword_3	newtag
1	Fantasy	Urban	Steampunk	Fiction:Fantasy-Urban-Steampunk
2	Romance	Vampire	NONE	Fiction:Romance-Paranormal
3	Romance	Suspense	NONE	Fiction:Romance-Suspense
4	Romance	Thriller	NONE	Fiction:Romance-Suspense
5	Romance	Werewolves	NONE	Fiction:Romance-Paranormal
6	Alpha Male	Romance	NONE	Fiction:Romance-Paranormal
7	Fantasy	Shapeshifter	NONE	Fantasy:Shapeshifter
8	Erótico	Romántico	Humor	Fiction:Romantic-Erotic-Humorous
9	testcases	suspense&thriller	supernatural	TEST:Testcases-Suspense&Thriller-Supernatural

Table **_tag_combination_rules** allows you to create new Work Tags whenever a combination of pre-existing Work Tag keywords exists. You own this table. New Work Tags are derived after all other Work Tag Scrubbing, including deriving new Work Tags from Comments. So, by the time this is used by Q&S, all other Work Tags have already been fully derived and scrubbed. This specific derivation is always done for all books, not just for selected books. So, even if you select just one (1) book to scrub, this derivation will automatically be done for all books in that particular Q&S library that have Work Tag keyword matches. Keyword <> Work Tag. A Work Tag of Fiction:Romance will satisfy ***tag_keyword_1*** in rules 2-5 in the picture above. The Work Tag keywords are NOT case sensitive. A few example seed entries in this table are added when it is first created by Q&S. You should delete any entries that you do not like, as you own it.

Work Tag Capitalization Rules Table

Table **_tag_capitalization_rules** is used automatically by Tag Scrubbing and Tag Minimization. Its purpose is to allow you to define how you want Work Tags that have complex punctuation to be capitalized.

“*Factual:History:War&MilitaryWorld War II-North Africa-Libya*” would normally be “*Factual:history:war&military-world war Ii-north Africa-libya*” unless the Tag Rules table had the latter as an “oldtag” and the former as the “newtag”. Even then it would be possible that the capitalization might turn out to be suboptimal. Hence, this new table.

The “rules” are only what you see, and they do what they say.

As delivered, these rules first make the Work Tags lower case (priority of 999 is the highest), then makes the very first letter upper case. Based on other matches, it either capitalizes or ensures that a letter stays lowercase. The lowercase rules at the bottom of the table are essentially superfluous because of the very top rule. They are provided as an example. Delete any or all of them.

Note these regexes have no leading or trailing “/”s.

The [#] sign examples allow you to turn “BLAH#BLAH” into “BLAH # BLAH”.

You can also delete all spaces with a new rule if you wish to add it.

This is your table now.

Table **_tag_capitalization_rules**

regex	rule	priority
[a-z]+	lowercase	999
^[a-z]	uppercase	950
^Fiction[:]	titlecase	900
^Factual[:]	titlecase	900
deletemenow	delete	200
[#]	addspace_left	200
[#]	addspace_right	200
[:][a-z]	uppercase	100
[-][a-z]	uppercase	100
[.][a-z]	uppercase	100
[&][a-z]	uppercase	100
[][a-z]	lowercase	100
[][a-z]	uppercase	100
[a-z][]	lowercase	100
[a-z][:]	lowercase	100
[a-z][-]	lowercase	100
[a-z]&]	lowercase	100
[a-z][]	lowercase	100
[a-z][.]	lowercase	100

Work Tag String Replacement Rules Table

This table is used automatically by Tag Scrubbing and Tag Minimization to change all pertinent Work Tags for all books every time those activities are executed, even if for only a single book. Its purpose is to give you granular control of arbitrary strings (phrases or partial words or whole words) that are found in Work Tags. This functionality **is** case-sensitive. The 'seed' values provided with this table are entirely related to proper Spanish spelling. Although the seed values happen to be all “words”, any partial word or phrase may be used.

One important caveat is best explained via a simple example:

The Spanish word for “Religion” is “Religión”. If a rule for Religion to be changed to Religión were created in a Q&S library that had both English and Spanish language books, the English Work Tags with the string “Religion” would be changed to “Religión”. For that reason, Religión and a few other Spanish words are not included as a seed entry for this table.

You own this table, and may safely delete any or all entries as you wish.

Table:

	id	old_string	new_string
	Fi...	Filter	Filter
1	9	Antropologia	Antropología
2	32	Autobiografia	Autobiografía
3	33	Belico	Bélico
4	31	Biografia	Biografía
5	28	Clasico	Clásico
6	29	Comunicacion	Comunicación
7	25	Divulgacion	Divulgación
8	24	Economia	Economía
9	21	Erotico	Erótico
10	20	Espana	España
11	1	Espanol	Español
12	18	Fantasia	Fantasia
13	2	Fantastico	Fantástico
14	17	Ficcion	Ficción
15	16	Filosofia	Filosofía
16	19	Investigacion	Investigación
17	8	Magico	Mágico
18	12	Poesia	Poesía
19	11	Policia	Policía
20	10	Política	Política
21	23	Psicologia	Psicología

Manga Awareness

If the string or sub-string of “manga” appears anywhere in the Author, Title, Series or Tags of a book, then “special Manga processing” is performed. The easiest way to invoke this is for your Manga books to all have a Tag of “manga”, or perhaps an Author of “mangaka”.

The primary benefit of this special processing is that it updates the Volume and/or Chapter numbers in the Work Series column. It also cleans up the titles to the extent they exist.

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	WorkTags	Tags
1	book_... ✓	mangaka	mangaka	Large Explosion	[ams]Large Explosion - Ch09.rar	Chapter [9]		Manga	manga
2	book_... ✓	mangaka	mangaka	Blue Manga	[Kissimy FI] Blue Manga [Chapter 001-700] [Complete] (Size: 2.44 GB)	Chapter 001-700 [1]		Manga	manga
3	book_... ✓	mangaka	mangaka	Tokyo No Ie Chapter 2	Tokyo no Ie chapter 2	Manga [2]		Manga	manga
4	book_... ✓	mangaka	mangaka	Chapter 02	Chapter 02	Manga [2]		Manga	manga
5	book_... ✓	mangaka	mangaka	Tokyo No Ie Volume 3	Tokyo no Ie volume 3	Manga [3]		Manga	manga
6	book_... ✓	mangaka	mangaka	Tokoyo No Ie 3	Tokoyo no Ie 3	Manga [3]		Manga	manga
7	book_... ✓	mangaka	mangaka	Claymore Volume 08	Claymore__Volume_08	Manga [8]		Manga	manga
8	book_... ✓	mangaka	mangaka	Japan 19	19.png	Manga [19]		Manga	manga
9	book_... ✓	mangaka	mangaka	Japan	v03_c07.rar	Manga_Volume_03 [7]		Manga	manga
10	book_... ✓	mangaka	mangaka	Alice Someone	[Shogun]_Alice_Someone_v03_c07.zip	Manga_Volume_03 [7]		Manga	manga
11	book_... ✓	mangaka	mangaka	Joe Blow	[Japan] Joe Blow v03 c07 blah blah blah	Manga_Volume_03 [7]		Manga	manga
12	book_... ✓	mangaka	mangaka	Alice Someone	[Shogun] Alice Someone volume 03 chapter 07	Manga_Volume_03 [7]		Manga	manga

Note: The above test data is 100% artificial (not “real”), and is copyright-free.

Manually Forcing Series Extraction from Title [Optional]

Optional. If for some reason you wish to import a large number of special books with both the title and the series in the title field, and you want Q&S to automatically relocate the metadata to the correct locations, you may use any of the templates shown below. Each line below would be a possible “Title” for each of the books, properly substituting the correct values for the placeholder words used in the templates below. The “**Series=**” is not a placeholder. The symbols, punctuation and spacing must be followed as shown in each choice. **Authorname** may be either FIRSTNAME LASTNAME or LASTNAME, FIRSTNAME. Initials should be in the format “J.R. Smith”, with no space between the initials. Testing first with only a few books is advised.

If **Authorname** is specified, you should verify that the Authors are in either the Global Author Table or the Pristine Author Table to enable detecting the **Authorname** and putting it where it belongs in the metadata. The Global Author Table comes with ~19,000 Authors preloaded. You should add “new” valid Authors to your Pristine Author Table. Actual table name: **_pristine_authors** . Always run an Author-Level Scrubbing job after updating that table, since it first does critical maintenance to both it and the Global Author Table.

Authorname - Someseriesname 02: Sometitle Here
Authorname - Sometitle Here: Someseriesname 02
Series=Someseriesname 01
title - Someseriesname, #2
title (Someseriesname, #2)
title [Someseriesname 2]
title [Someseriesname, #2]
title, Someseriesname 2
title; Someseriesname 2
title: Someseriesname, #2

Deriving User Custom Columns from Title

This capability allows you to update your own non-Q&S custom columns directly from the Title/Work Title.

Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	SomeCCName	SomeCCName2	WorkTags	Tags
1	book_ok ✓	Aaron Blow	Aaron Blow	The Boy Nazi	Dark Evil 11: The Boy Nazi {{20}:Something 1}	Dark Evil [11]		Something 1		Vampires, Fiction:Westerns	FIC033000, Vampires
2	book_ok ✓	Aaron Blow	Aaron Blow	The Boy Nazi	Dark Evil 11: The Boy Nazi [21]:Something 2}	Dark Evil [11]		Something 2		Vampires, Fiction:Westerns	FIC033000, Vampires

Assume for this example that Custom Columns 20 and 21 were created for some user-defined purpose.

To activate CC Derivation in Q&S from the Work Title, you must add entries into Table `_title_cc_control`. The entries tell Q&S (a) What the CC id number is (e.g. 20); (b) the type of the CC (only text and comments are valid for use by Q&S for this purpose); (c) which Q&S derivation rules are valid for a particular CC id (only rules 1 and 2 are currently valid); and, (c) which tells Q&S whether or not your entry is "active". You can create rules for future use and mark them inactive by assigning a 0 to the active column. One `cc_number` can have as many rules as are valid (currently Only 2).

For your data protection, you should configure only the combinations that you intend to use. Also, only `cc_numbers` greater than 19 are allowed since the official Q&S library that is delivered uses 19 and below. Any new custom columns will necessarily be above 19. Do not attempt to create any new custom column with the same label as any preexisting Q&S custom column, such as "status".

TABLE <code>_title_cc_control</code>			
<code>cc_number</code>	<code>datatype</code>	<code>rule</code>	<code>active</code>
20	text	1	1
21	comments	1	1
21	comments	2	1

id	label	name	datatype	mark_for_delete	editable	display	is_multiple	normalized
20	some_cc_name	SomeCCName	text	0	1	{"description": "A ...	0	1
21	someccname2	SomeCCName2	comments	0	1	{"description": "lo...	0	0

To view these tables, use the same tool that you use to maintain the other Q&S rules tables, such as DB Browser for SQLite from <http://sqlitebrowser.org/>

Currently there are only 2 valid values for Rule in table `_title_cc_control`: 1 and 2.

Rule 1: The pattern of the to-be-derived data must be exactly: `{{NN}:xxxxxxxxxxxxx }` where NN is the `cc_number` (which is also the CC id), and xxxxxxxxxxxxx is whatever you want to go into custom column NN.

Rule 2: The pattern of the to-be-derived data must be exactly: `[NN]:xxxxxxxxxxxxx` where NN is the `cc_number` (which is also the CC id), and xxxxxxxxxxxxx is whatever you want to go into custom column NN.

These patterns were selected so that they would not interfere with existing Work Series derivation patterns already being used by Q&S.

Use whichever you prefer, or both. They work the same. *Don't forget the } or the] at the end of the patterns.* That tells Q&S where your xxxxxxxxxxxxx ends.

These patterns can be used anywhere within the Title (which is copied to Work Title): at the beginning; in the middle; at the end. They will be removed from the Work Title after derivation.



Quarantine & Scrub User Guide

Dewey Decimal/Library of Congress

Classification per <http://www.oclc.org> Using ISBN or ISSN
Add DDC and/or LCC Via New Custom Columns You Create

n.b. For an “Industrial Strength” DDC/LCC Plug-In, download and install “Library Codes”.

Customize QuarantineAndScrub

Paths

Enter the Full Path of Your Pristine Calibre Library, Including metadata.db

S:/Users/Public/Public eBooks/CalibreMain/metadata.db

Q&S Library with the Best Tag & Title Rules and Web Source Series Validation Data (WSSVD)

S:/Users/Public/Public eBooks/QuarantineAndScrub/metadata.db

Q&S Target Library # [1][2][3][4] for Tag & Title Rules and WSSVD

S:/Users/Public/Public eBooks/QuarantineAndScrub_prescrub/metadata.db

S:/Users/Public/Public eBooks/QuarantineAndScrub_postscrub/metadata.db

none

none

Classify: DDC & LCC [Selected Books]

On the Main Menu

Miscellany

Enter the Maximum Number of Work Tags to Allow Per Book via Tag Priorities

1

Q&S Derive Genres: Custom Column Lookup #Name to Update

#genre

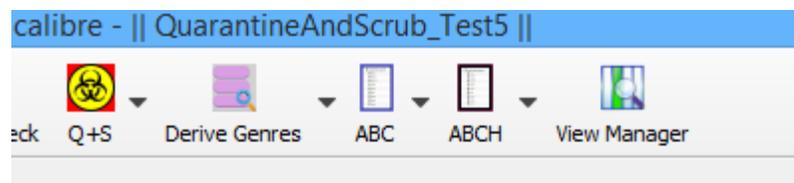
DDC and/or LCC: Custom Column Lookup #Names

#ddc *

#lcc *

Activate Classification?

OK Cancel



DDC	LCC	OCLC-OWI	ISBN
813.6	PS3601.A44	694994249	9781445871059
813.6	PS3610.A4475	192151149	9781605044040
940.48	D568.7	1922422	9781421800004

Edit a custom column

Lookup name owi

Column heading OCLC-OWI

Column type Column built from other columns Show checkmarks

Description OWI is an identifier at oclc.org which is used to obtain DDC and LCC

Template {identifiers:select(oclc-owi)} Default: (nothing)

Sort/search column by Text Show in tags browser Show as HTML in book details

OCLC-OWI is always added as a Calibre Identifier if DDC or LCC is updated. If you want to view it as a clickable URL link in the Book Details, create a hidden custom column as shown here. Note that the Template is:

{identifiers:select(oclc-owi)}



Author Renaming

Change an Author Name into Another Author Name for Any Reason You Wish: Standardization of Spelling; Pseudonym to Real Name; Pseudonym to Other Pseudonym; Real Name to Pseudonym.

Status	Freeze	WorkAuthor	*	Author(s)	W
1	book_ok ✓	Dean Koontz	*	Anthony North	Test Author Mapping T
2	book_ok ✓	Gore Vidal	*	Edgar Box	Test Author Mapping T

Quarantine And Scrub: Author-Level Scrubbing

```

Starting job: Quarantine And Scrub: Author-Level Scrubbing
Library DB: S:/Calibre/Q&S/QuarantineAndScrub_Test4/metadata.db
Tue Mar 17 09:35:59 2015
Python: Windows CPython 2.7.9
SQLite Version: 3.8.4
SQLite DB Locking Mode: EXCLUSIVE [GUI Is Locked]

=====
Beginning Author Level Scrubbing
=====

For Your Reference
=====
Number of Global Authors: [16694]
Number of Global Series: [5201]
Number of Pristine Authors: [27]
Number of Pristine Series: [0]
=====

Number of books: [2]
Number of books with a Work Author: [2]
Number of books with a Status of 'book_ok': [0]
Number of books with a Status of 'dirty': [2]

'New' Valid Real Authors Must Be Added to the Pristine Author Table Prior to Author Level Scrubbing.
'New' Valid Real Authors Are Not Already Global Authors and Not Already Pristine Authors.
Calibre>Q&S>[Menu] Q&S Copy Work to Real > Copy Valid Real Authors to Q&S Pristine Author Validation Table [Selected Books]
Be certain to select only truly valid Real Authors to copy to the Q&S Pristine Author Validation Table.

Number of Author Names Changed Per Author Mapping Rules: [2]
```

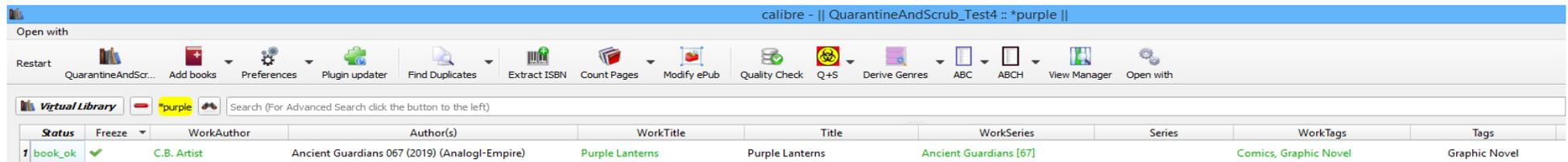
TABLE _author_mapping Search Show All

id	name	newname
1	E.E. "Doc" Smith	E.E. Smith
2	Acton Bell	Anne Brontë
3	Aaron Wolfe	Dean Koontz
4	Alice Addertongue	Benjamin Franklin
5	Amanda Cross	Carolyn Gold Heilbrun
6	Anne Rampling	Anne Rice
7	A.N. Roquelaure	Anne Rice
8	* Anthony North	Dean Koontz
9	Clive Hamilton	C.S. Lewis
10	David Axton	Dean Koontz
11	Deanna Dwyer	Dean Koontz
12	* Edgar Box	Gore Vidal
13	J.D. Robb	Nora Roberts

Author-Level Scrubbing automatically applies these rules whenever it is run.

Graphic Novel/CBZ/CBR Scrubbing

Special Processing for Scrubbing of Author and Title Invoked When: format is CBZ or CBR or any other format has a Tag of "Graphic Novel". A log message will say so. Tables **_title_rules** and **_series_rules** have special capitalization rules for CB characters; add more as you feel the need. An Author of "Unknown" is changed to "C.B. Artist", and a new Work Tag of "Comics" is automatically added. Using tables **_author_mapping** and **_tag_rules**, you may change those values to whatever you please.



Status	Freeze	WorkAuthor	Author(s)	WorkTitle	Title	WorkSeries	Series	WorkTags	Tags
book_ok	✓	C.B. Artist	Ancient Guardians 067 (2019) (Analogl-Empire)	Purple Lanterns	Purple Lanterns	Ancient Guardians [67]		Comics, Graphic Novel	Graphic Novel

Periodicals: Scrubbing

Periodicals with a Title with a pattern such as "Some Name - March 22, 2015 XXX" or "Something - March 2015 XXX" will **not** be scrubbed or changed whatsoever. A log message will say so.

Derive Work Tags from Book Awards Table

Book-Level Scrubbing and also standalone Tag Scrubbing will create a new Work Tag using the Book Awards Table, **_book_awards**. Because of slight differences in Title and Author names, Q&S calculates a "% Similarity" of each book's Work Author and Work Title to the Book Awards Table Author and Title. If each is ≥ 0.85 , the Award is created as a new Work Tag. This is done prior to any reference to the Tag Rules table, so you may create a renaming rule to change the official Award name to whatever you wish. Example: "Nebula Award" could be mapped to "Nebula Award for Science Fiction" if you wished.

TABLE

id	author	title	award
17958	David Marusek	The Wedding Album	Nebula Award
18734	David McCullough	John Adams	Pulitzer Prize
17840	David McCullough	Mornings on Horseback	National Book Award
18816	David McCullough	Mornings on Horseback	Pulitzer Prize
18763	David McCullough	Truman	Pulitzer Prize
17536	David Mitchell	Black Swan Green	Man Booker Prize
17342	David Mitchell	Cloud Atlas	Man Booker Prize
18185	David Mitchell	Cloud Atlas	Nebula Award
17398	David Mitchell	The Bone Clocks	Man Booker Prize
17587	David Mitchell	The Thousand Autumns of Jacob de Zoet	Man Booker Prize
16780	David Moles	Finisterra	Hugo Award
15241	David Morrell	Creepers	Bram Stoker Award
15226	David Niall Wilson	The Gentle Brush of Wings	Bram Stoker Award
17405	David Nicholls	Us	Man Booker Prize
17163	David Park	The Light of Amsterdam	International Impact Dublin Literary Award

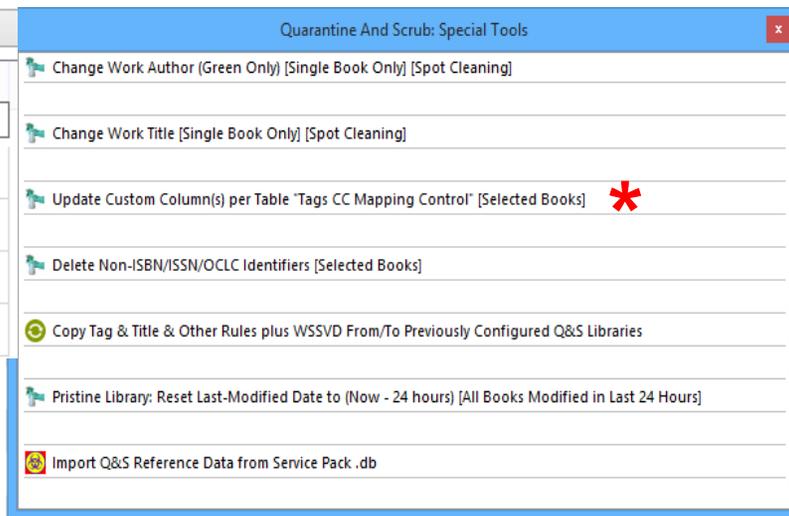
Derive Custom Columns from Work Tags

By configuring Table **_tags_cc_mapping_control**, and running its job after Book-Level Scrubbing has completed, you can automatically populate multiple custom columns with either a Work Tag or any arbitrary value that you specified based on a Work Tag used in a rule in the **_tags_cc_mapping_control** table.

tag	cc_number	datatype	rule	cc_value	active
Filter	Filter	Filter	Filter	Filter	Filter
Fiction:Dystopian	20	text	T1	NULL	1
Fiction:Dystopian-Other	20	text	V1	output_tag_rule_v1	0
Fiction:Science	21	comments	T2	NULL	1
Fiction:Science-Space Opera	21	comments	V2	output_tag_rule_v2	0

Rules Defined Below

Active = 1: Job will use the rule



CC 20 CC 21

SomeCCName	SomeCCName2	DDC	LCC	ISBN	OCLC-OWI	ISSN	WorkTags
output_tag_rule_v1	Fiction:Science			9780684843285	<a href="..."		Fiction:Dystopian-Other, Fiction:Other, Fiction:Science, Fiction:Science-Space Opera, Need5Fixing
Fiction:Dystopian	Fiction:Science				<a href="..."		Fiction:Dystopian, Fiction:Dystopian-Other, Fiction:Science, Fiction:Science-Space Opera
Fiction:Dystopian	Fiction:Science	641.5	TX651	9780684843285	<a href="..."	0047-2689	Fiction:Dystopian, Fiction:Dystopian-Other, Fiction:Science, Fiction:Science-Space Opera, Fiction:Su...
Fiction:Dystopian	Fiction:Science	641.5	TX651	9780684843285	<a href="..."		Fiction:Dystopian, Fiction:Dystopian-Other, Fiction:Fantasy-Urban-Steam-punk, Fiction:Science, Ficti...

- * **Rule T1:** use the Work Tag in the table as the Value for the custom column, and leave the Work Tag alone
- * **Rule T2:** use the Work Tag in the table as the Value for the custom column, but then delete the Work Tag
- * **Rule V1:** use the CC Value in the table as the Value for the custom column, and leave the Work Tag alone
- * **Rule V2:** use the CC Value in the table as the Value for the custom column, but then delete the Work Tag

Rules for a specific Work Tag are applied in this sequence: **V2,V1,T2,T1**. So, multiple rules for same Work Tag for same CC can exist.



Derive a Pseudonyms Custom Column from Global Pseudonyms

If desired, a User Custom Column may be derived by mapping a Real Author to the Author in the Global Pseudonyms Table, **_global_pseudonyms**. The images below explain how to do this. "Quick Create" by selecting "People's Names".

Table: **_global_pseudonyms**

Which uses this table

	pseudonym	author	Author(s)	Pseudonyms
	Filter	Filter		
1	A.A. Craig	Poul Anderson		
2	A.A. Fair	Erle Stanley Gardner		
3	Aapeli	Simo Puupponen		
4	Aaron Wolfe	Dean Koontz		
5	Abigail Van Buren	Jeanne Phillips	Andre Norton	Andrew Norton, Andy Norton, Norton Andrew
6	Abigail Van Buren	Pauline Phillips		
7	Abram Tertz	Andrei Sinyavsky		
8	Abu Nuwas	Hasin Ibn Hani Al Hakami	Angie Sage	Anne Thyme, Annie Time, Jules Rosemary

And Updates the CC

```
Starting job: Quarantine And Scrub: Miscellany - Update Pseudonym Custom Column
Library DB: S:/Calibre/Q&S/QuarantineAndScrub_Test6/metadata.db
Mon Jun 01 12:39:00 2015
Python: Windows CPython 2.7.9
SQLite Version: 3.8.4
SQLite DB Locking Mode: EXCLUSIVE [GUI Is Locked]

Beginning Miscellaneous Scrubbing
Update Pseudonym Custom Column [Selected Books]

-----
Q&S Configured Pseudonym Custom Column #Name to be used is: #pseudonyms
-----

This job uses Real Author, not Work Author.
---- An Author May Have Many Pseudonyms----
New Pseudonym(s) for Author and Book: Fiat Lux - 3291 - Trusty Light, genëthëtö phös
New Pseudonym(s) for Author and Book: Fiat Lux - 3292 - Trusty Light, genëthëtö phös
New Pseudonym(s) for Author and Book: Fiat Lux - 3293 - Trusty Light, genëthëtö phös
New Pseudonym(s) for Author and Book: Bella Andre - 3294 - Andys Beauty
New Pseudonym(s) for Author and Book: Bella Andre - 3295 - Andys Beauty
New Pseudonym(s) for Author and Book: Bella Andre - 3296 - Andys Beauty
New Pseudonym(s) for Author and Book: Andre Norton - 3301 - Andrew Norton, Andy Norton, Norton Andrew
New Pseudonym(s) for Author and Book: Angie Sage - 3304 - Anne Thyme, Annie Time, Jules Rosemary
```

Review the Job Log



FAQs

1) Why is only one of several Real Authors copied into #Work Author?

Answer: Q&S is for scrubbing trashed-out metadata, not clean metadata, so it must handle "Authors" like: "Georgina Gentry - Iron Knife's Family 01 - Cheyenne Captive"

2) What if my Real Authors are perfect? I want to keep them.

Answer: Easy. When it comes time to copy Work to Real, do not copy Work Author to Real Authors for that book.

3) After Author-Level Scrubbing runs, only some (or even none) of the books have a status of "auth ok". Is that correct?

Answer: Yes. It only sets a new status for books with Authors that it changes.

4) I ran Author-Level Scrubbing, but most or all of the books are still "dirty". Is this right?

Answer: Yes. See #3, above.

5) Some of my books that I am scrubbing are "perfect". Nevertheless, Q&S changes them. Why?

Answer: Q&S is for scrubbing trashed-out metadata, not clean metadata. Its presumption is that it is all dirty. Do not scrub "perfect" books. Or, see #2, above, but apply the concept to all of the Work Data, not just Work Author.

6) I have 30,000 books for which I have downloaded metadata, but the metadata is still a mess and requires a lot of manual labor to fix. 30,000 books would take ~ 30 hours to copy from my Calibre library to a new Calibre QuarantineAndScrub library. Is there a way to reduce the time from 30 hours to something much less?

Answer. Yes. You can reduce the 30 hours to 10 minutes by following the attachment to the Original Post named: "*For Technical Users Only Metadata db Swap Instructions.txt*". Also read its companion .txt file regarding Triggers.