

# Title Normalization & Quality Control: Making Excel Work for You and Not the Other Way Around

2019 UMD Libraries Research & Innovative Practice Forum

Claralyn Burt, Graduate Assistant, CRDM

## Introduction

Every year, the Continuing Resources & Database Management Department (CRDM) systematically renews the journal and other serial subscriptions for the UMD libraries. This process involves obtaining title lists from each publisher and comparing them to the order history from Aleph. This is important, because publishers buy and sell titles. Knowing which titles are new to a publisher's package or have left a package influences how and via which platform a library user will access the resource. To accomplish this, an initial comparison using a series of vLookUp functions is made between the title lists and the Aleph order file. The CRDM department then has the raw data to make informed decisions about journals in the upcoming year. While it sounds simple, there are many moving parts resulting in a multi-step, multi-month workflow. This past 2018 renewal season, I streamlined the process in two innovative ways: [title normalization](#) and using the [IF function in Excel as a quality control step](#).

## Title Normalization

I began by cleaning the data to normalize the titles. I tried to have it match the Aleph Data as much as possible in formatting and syntax. Each step is listed below.

Remove the excess information from the Title cell using "Find + Replace." This includes words like "- ONLINE" or "SUPPLEMENT."

Copy the trimmed Title and "Paste (Values Only)" so the formulas aren't transferred. This becomes the "clean" title to work with.

Using the TRIM function, delete the excess spaces that may have crept in.

Using VLOOKUP with the ILS as a match point, pull in the title from the Aleph Order History.

Following similar steps used for the Original Title, I cleaned up the Aleph Title by trimming excess spaces, standardizing capitalization, and Pasting Values Only to make a "clean" title.

## Using the IF Function for Quality Control

After completing the rigorous vLookUp function, I added a quality control step with an IF function to ensure that the matches were accurate. It asked if two cells were the same. For example, =IF(Ax=Bx,"ok","not ok"). However, I realized this was not a wide enough formula. It did not account for determiners such as "the." So I created a new formula:

**=IF(Ax=Bx,"ok",IF("The "&Ax=Bx,"ok",IF(Ax="The "&Bx,"ok","not ok")))**

It essentially added nested syntax formulas by combining "The" with the contents of the specific cell. Nested formulas allowed for flexibility. For example, depending on the title list, some aggregators left out "the" as a determiner. I could insert whichever singular nested formula I needed for that specific list, but the principle of the formula remained the same.

A version of this formula can be seen in practice below:

K	L	M	N	O	P	Q
imme	TITLE_clean	v_Title	v_Title_clean_1	v_Title_clean_2	v_TITLE_clean	match?
M(J2)	Discourse & Society	=VLOOKUP(E2,Aleph_Discourse & society	Discourse & Society	Discourse & Society	=TRIM(O2)	=IF(L2=P2,"ok",IF("The "&L2=P2,"ok","not ok"))
M(J3)	Psychology Of Music	=VLOOKUP(E3,Aleph_Psychology of music	Psychology Of Music	Psychology Of Music	=TRIM(O3)	=IF(L3=P3,"ok",IF("The "&L3=P3,"ok","not ok"))
M(J4)	Psychological Science	=VLOOKUP(E4,Aleph_Psychological science	Psychological Science	Psychological Science	=TRIM(O4)	=IF(L4=P4,"ok",IF("The "&L4=P4,"ok","not ok"))
M(J5)	Simulation : Transactions Of The Soci	=VLOOKUP(E5,Aleph_Simulation	Simulation	Simulation	=TRIM(O5)	ok
M(J6)	Sociological Review - England	=VLOOKUP(E6,Aleph_The sociological review	The Sociological Review	The Sociological Review	=TRIM(O6)	ok

**=IF(L2=P2,"ok",IF("The "&L2=P2,"ok","not ok"))**

## Conclusion

I added two new innovations to the workflow of comparing serial title lists with the Aleph order history. I cleaned the raw data so it was uniform with the formatting of the Aleph order history. Because I took the time to standardize and normalize my data at the beginning of this process, I could create a better, more inclusive formula. I used these steps with 11 different publisher packages. Ultimately, I did not have to manually check hundreds of titles because I engineered Excel to do it for me. This saved time and money in meeting the needs of the users and researchers across the seven University Libraries.