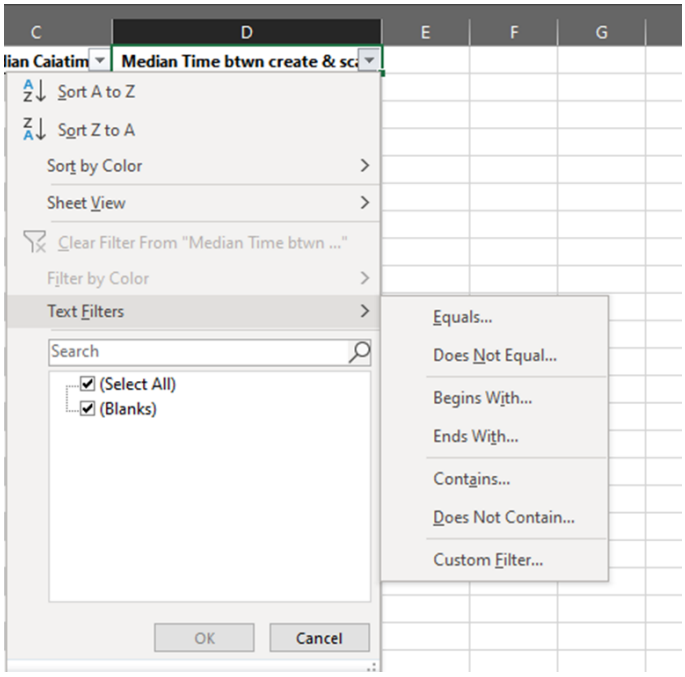
Instructions for APC Sources Data

In this document, we will walk you through some tips and tricks for using APC Sources Data to 1. Filter out the corresponding author data that matches your institution, 2. Clean up names, and 3. Save yourself some time while looking up e-mail addresses. Please consult our Sample Data Template for illustrations of how these items work.

Step 1: Find the corresponding authors affiliated with your institution.

Unfortunately, the data can contain multiple corresponding authors. You want the ones affiliated with your institution. To do this work:

1. Open the data in Excel.
2. Convert the data to a table. (I copied the data into a new sheet, on the Home tab clicked "Format as table" and selected that my data had headers.
3. Now you should be able to select a column header and select text filters, then "Contains".
   a. 

4. Enter your search term. (You can use University name for the search term, or a version of it.)
5. Wait while Excel processes it (this can take a little bit of time).
6. After this step, you should have a filtered dataset with just rows that match your query. You can then copy that filtered dataset into a new, clean s
7. The Corresponding author column may include more than one author affiliated with your institution. You may need to manually check the authors – for example, selecting faculty over students or authors that have shown up in the spreadsheet less frequently over authors with multiple articles in the datasets, ultimately leaving only one corresponding author from your institution in this field.

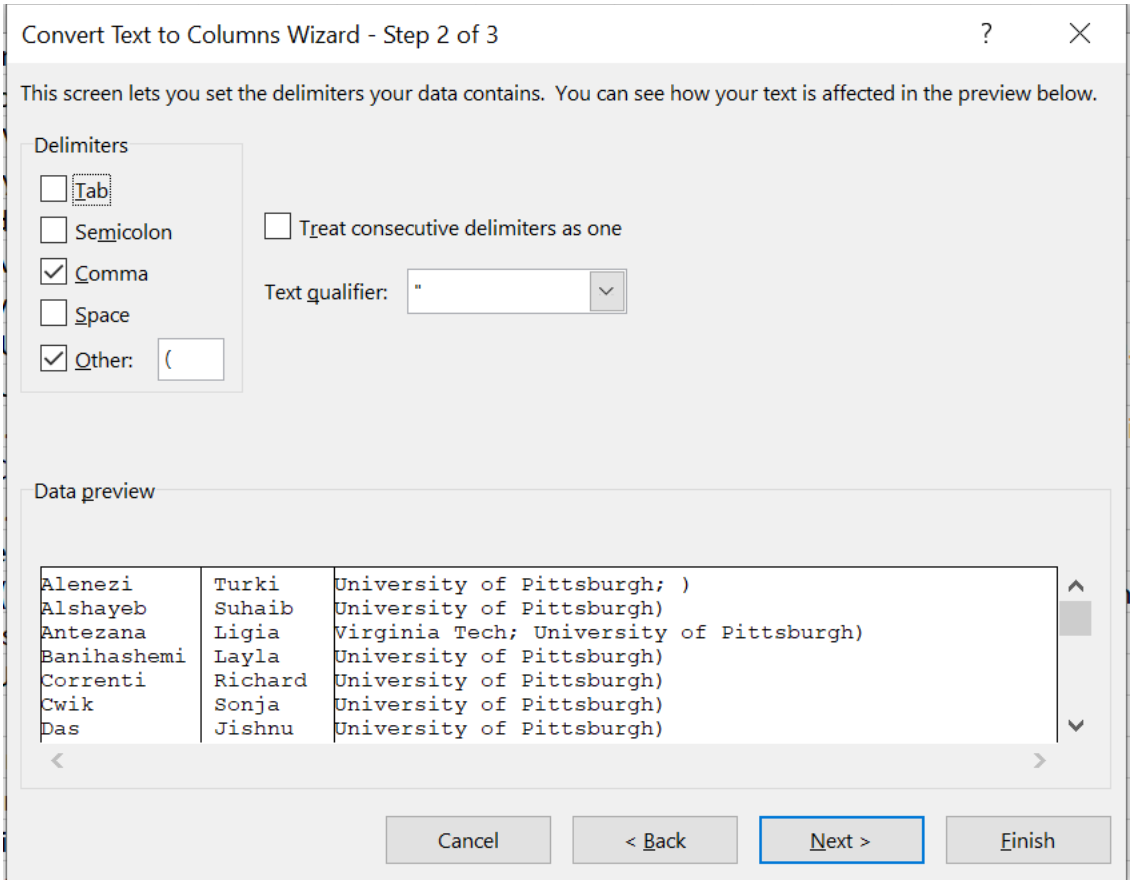
Step 2: Use Text-to-Columns and CONCAT to create full names

The data export usually contains name data in this format: Lastname, Firstname

When you do a mail merge, this format is difficult to work with. The existence of commas can also interrupt other searching and filtering that we want to do in future steps. As such, you want to transform your data into Firstname Lastname.

In our template tab “Name Data”, you can see how to do this using the CONCAT function in Excel.

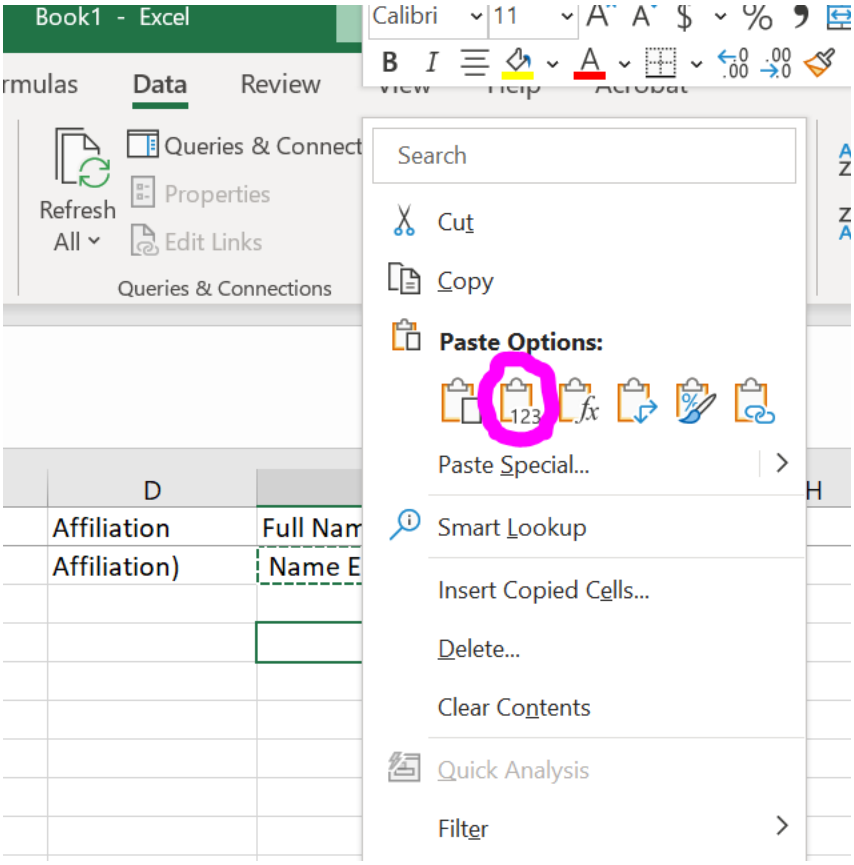
In this tab “Name Data”, copy the corresponding author names from the Corresponding Authors field. Then select the column, and in Excel go to the Data tab and select “Text to Columns”. Select “Delimited,” then these options. Make sure to click “Next” (not Finish); on the next screen you’ll get the option of what cells to use as the output for this delimiting, otherwise it will replace your current column.



You’ll get the names and affiliations separated into three columns: last name, first name, and then the affiliation.

If we want the two names to be put together (e.g. “Lauren Collister” in one field) we can use =CONCAT() to do that in a new column. In the sample spreadsheet, we use =CONCAT(C2,B2) to concatenate the first name before the last name.

Copy the column with Full Name into your working spreadsheet by copying the column, then right clicking and using the “Paste Values” option (circled in pink on the screenshot below).



STEP 3: Create our List of Full Names, Type, and Discipline to create an e-mail XLOOKUP database.

In this step, we will prepare to save Future You time and energy by not duplicating your work looking up the same e-mail addresses over and over again. Since many author names may appear on multiple lists (perhaps they publish in different disciplines, or have some hybrid and some gold), this step can save future you some time.

First, start by looking up e-mails for one of your sets of data. Let's say you start with Gold HSS. Once you have your e-mail addresses for those authors, you can begin to build a tab in your master data spreadsheet that functions like a database.

Once you have these data, create a new tab in your spreadsheet and name it "All Emails". You can use the Sample Data tab as an example. In this you will want columns for: Email, Name, Type, Discipline.

(Note: Use the Full Name that we created in Step 2; the original corresponding author name data contain commas that can be very annoying for the functions we want to use later.)

Copy your E-mail and Name columns from your original data into the new All Emails tab into their respective columns. At this point, make sure to add the data for type and discipline. (Pro tip: You can

type one value into one cell at the top or bottom of this new list, e.g. HSS, and then copy it, select the entire rest of the cells next to the data you just pasted, and paste that value (HSS) into those columns.) Use the column labeled “Fields of Research Codes” and the following divisions of those codes to assign disciplines.

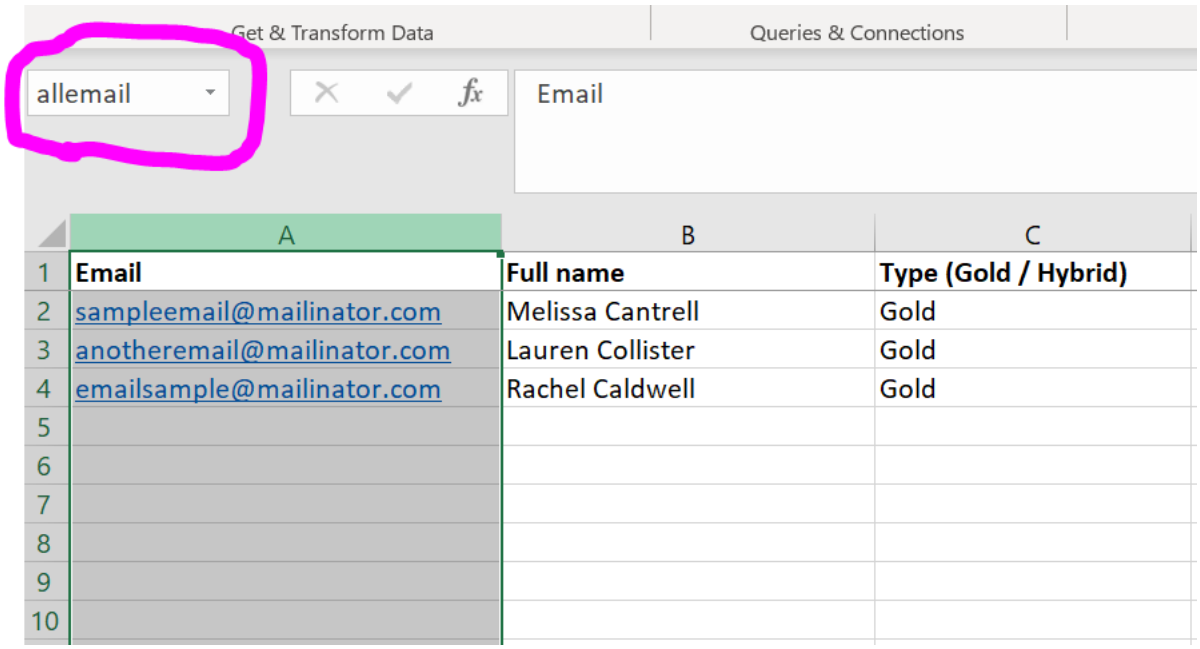
DIVISION		
30	AGRICULTURAL, VETERINARY AND FOOD SCIENCES **with exception below**	NSE
30 09	veterinary science	HM
31	BIOLOGICAL SCIENCES	NSE
32	BIOMEDICAL AND CLINICAL SCIENCES	HM
33	BUILT ENVIRONMENT AND DESIGN	NSE
34	CHEMICAL SCIENCES	NSE
35	COMMERCE, MANAGEMENT, TOURISM AND SERVICES	HSS
36	CREATIVE ARTS AND WRITING	HSS
37	EARTH SCIENCES	NSE
38	ECONOMICS	HSS
39	EDUCATION	HSS
40	ENGINEERING	NSE
41	ENVIRONMENTAL SCIENCES	NSE
42	HEALTH SCIENCES	HM
43	HISTORY, HERITAGE AND ARCHAEOLOGY	HSS
44	HUMAN SOCIETY	HSS
45	INDIGENOUS STUDIES	HSS
46	INFORMATION AND COMPUTING SCIENCES	NSE
47	LANGUAGE, COMMUNICATION AND CULTURE	HSS
48	LAW AND LEGAL STUDIES	HSS
49	MATHEMATICAL SCIENCES	NSE
50	PHILOSOPHY AND RELIGIOUS STUDIES	HSS
51	PHYSICAL SCIENCES	NSE
52	PSYCHOLOGY **with exception below**	HM
52 05	social and personality psychology	HSS

In addition to these basic divisions for Field of Research codes, you may also notice that some publication have multiple, more granular codes. If there is a Field of Research code anywhere in the string that starts with 32 or 42, put that publication into the HM disciplinary bucket. If there is a string of Field of Research codes that conflicts with the first code, then you may need to consider a manual review for those outliers. What this manual process looks like will differ from one project to another, and you may wish to consider other criteria in the Dimensions data such as journal title or author department to determine best disciplinary fit.

[Provide a couple of examples]

As a heuristic for this research project, the investigators kept publications within the original disciplinary bucket based on the first Field of Research code where there was ambiguity about fit, and only moved those publications where the first Field of Research code was a clear contradiction to disciplinary fit.

Now you have the beginnings of your database for lookup. Important!! Select all of Column A, where you have your e-mails, and LABEL THAT COLUMN “allemail”. You can label the column using the label to the left of text entry, as circled in pink in this screenshot. That column was previously just labeled “A” and now it is called “allemail”. Do the same thing for Column B, and call it “allfullname”. This is crucial for making the lookup function that we will describe in the next section.



	A	B	C
1	**Email**	**Full name**	**Type (Gold / Hybrid)**
2	sampleemail@mailinator.com	Melissa Cantrell	Gold
3	anotheremail@mailinator.com	Lauren Collister	Gold
4	emailsample@mailinator.com	Rachel Caldwell	Gold
5			
6			
7			
8			
9			
10			

Begin to work on another spreadsheet tab, for example Hybrid HSS. In the e-mail column in Row 2, paste in this lookup function:

=XLOOKUP(B2,allfullname,allemail)

What is this doing? It’s saying to Excel:

> Hey! Look this up for me. I’ve got this name in B2, “Melissa Cantrell”. Go back to that column that we labeled “allfullname” and see if “Melissa Cantrell” appears there. If it does, check to see if there’s a value in the corresponding row in “allemail” and put that e-mail address in A2 this cell where I’m typing this function.

If you copy this formula in that whole row, Excel will understand that you want to look up the Full Name value in the cell next to that formula, then search that other column in the other tab to see if there’s a matching name there with an e-mail address, and bring that e-mail address over to this column.

If there is an e-mail address, great! Excel will put it in your sheet so you don’t have to look it up again. If there is not e-mail address, Excel will do this:

	A	B	C	D
1	**Email**	**Full Name**	**Type**	**Discipline**
2	sampleemail@mailinator.com	Melissa Cantrell	Hybrid	HSS
3	#N/A	Jennifer Mezick	Hybrid	HSS
4	anotheremail@mailinator.com	Lauren Collister	Hybrid	HSS

That #N/A means that there was no e-mail address, because that author name – Jennifer Mezick – did not appear in our list of previously-gathered e-mail addresses. This tells you that you need to look up that e-mail address, type it into this spreadsheet.

Once you are done with your Hybrid HSS spreadsheet, copy and paste (remember to use the "Values" version of Paste) the e-mail addresses, names, Type, and Discipline over to your "All Emails" tab, making an even larger database of e-mails for XLOOKUP to consult. This way, if Jennifer Mezick also published in a Hybrid MED journal, XLOOKUP will find the corresponding e-mail address and save you from having to look it up again.

As a bonus, once you are done with all of your looking up, the "All Emails" tab is now a nice, clean version of your data that you can use for your mail merge to send out your e-mails.

Corresponding authors with multiple publications:

For those who have multiple publications for which they are corresponding author:

i. Send no more than two to each author

ii. If author has mix of gold and hybrid, choose most recent of each

iii. If author has only gold or only hybrid but within separate disciplines, pick most recent two from separate discipline

iv. If all pubs are the same model, same discipline, just choose the most recent two.

v. Use same email template but try to make publication prominent for differentiation (bold, top of email)

vi. For pubs that are removed (beyond the two selected), move to a separate tab but keep for the time being.