

OPA Excel Tips: Splitting cell values into columns

Excel recognizes comma separated value (csv) files and automatically places the values in each row into separate columns, however some file formats and data fields use different characters, or a set number of characters to split the file up into different variables.

The 'Text to Columns' function within Excel is very useful for converting a dataset consisting of a single column into multiple variables, and for splitting single variables. Another option for the latter is to use the LEFT and RIGHT functions (see separate OPA Excel Tip).

Example 1: Splitting PI names into separate columns.

In this example the Transplantation dataset is used. The data is already in columns but the PI column contains multiple PIs. The 'Text to Columns' tool can be used to split the column into individual PIs.

The screenshot shows an Excel spreadsheet with a column labeled 'PI Name(s) All'. A context menu is open over this column, and the 'Cut' option is highlighted. A red box surrounds the 'Cut' option and the text: "When splitting a column, Excel spreads contents across following columns. To avoid over-writing existing data, move column to be split to end of dataset. Here we cut and paste but as splitting the column splits the original cell, it is good practice to copy the column." Another red box contains the instruction: "(1) Right click at the top of the column to split and select 'Cut'".

(2) Move to the end of the dataset, highlight the column by right clicking at the top of it and select 'Insert Cut Cells'

(3) Right click on column J to split and select 'Text to Columns' in the 'Data' menu.

Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Fixed Width. This is correct, choose Next, or choose the data type that best describes your data.

Original data type: Delimited Fixed width

Preview of selected data:

PI Name(s) All
MESSENY, LOUISE MICHAEL
LIMAYE, KENNETH BEB
FERRE, JOHN P
RAO, WINSTON W

(4) The default is to split based on a 'Fixed width', a set number of characters. Instead select 'Delimited' meaning split cells based on a character (or characters). Then click 'Next'.

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters: Tab Semicolon Comma Space Other

Data preview:

PI Name(s) All
MESSENY, LOUISE MICHAEL
LIMAYE, KENNETH BEB
FERRE, JOHN P
RAO, WINSTON W

(5) Select to split by 'Semicolon' to split the column into individual PIs. Then select 'Next'.

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters: Tab

Semicolon Text consecutive delimiters as one

Comma Text qualifier: ' ' Text to be enclosed in

Space Other: _____

Data preview:

PI Name(s) All
 MESINA, LOUIS MICHAEL
 MARGULIES, KENNETH BER
 FISHER, JOHN P
 KAO, WINSTON W
 ...

Buttons: Cancel, Back, Finish

(6) Leave the formatting as 'General' and click on 'Finish'.

Microsoft Excel spreadsheet showing the result of the text-to-columns conversion. The 'PI Name(s) All' column has been split into two columns: 'PI' and 'Name(s) All'. The first row shows 'MESINA, LOUIS' in the 'PI' column and 'MICHAEL' in the 'Name(s) All' column. A red box highlights the 'Finish' button in the previous screenshot.

(7) The column will be split, with the first name in the original column and names added to later columns where they exist.

transplantation dataset - Excel

Perkins, Matt (NH400) [C] -

ApprId	Type	Actv	Project	FY	IC	Title	Abstract	SA	Text	Awd Tot \$	PI Name(s)	All	PI2	PI3	PI4	PI5
1	7957790	2	R01	HL075383	2010	HL	Mesenchy	DESCR	1	\$pe	\$411,250	MESINA, LOUIS MICHAEL				
2	8456490	1	R21	HL137777	2013	HL	Exercise a	DESCR	A	\$pec	\$200,000	LIBONATI, JOSEPH ROCCO [contact]	MARGUILES, KENNETH BER			
3	8245505	1	R01	AR041460	2011	AR	Applicati	DESCR	II	\$PECI	\$355,245	FISHER, JOHN P.				
4	8189876	1	R01	EY021766	2011	EY	Cell Ther	DESCR	PJ	Kao	\$530,406	KAO, WINSTON W				
5	8402215	1	R01	DK095001	2012	DK	The role o	DESCR	SPECIFC	\$333,825	METHWE, ALEXANDER					
6	8415397	1	U18	TR005364	2012	TR	Modeling	DESCR	SPECIFC	\$375,600	LYNCH, JOHN P.					
7	7472945	1	R43	DK083832	2009	DK	A perfluor	DESCR	1	\$SPECI	\$100,000	POO, RAMON E.				
8	8108873	1	R01	HL103709	2011	HL	Bioprocee	DESCR	1	\$SPECI	\$379,711	IZANAKAKIS, EMMANOUIL				
9	8545413	1	R01	DK097807	2012	DK	Culturally	DESCR	2	\$SPECI	\$211,402	BUCHWALD, DEBRA S				
10	8508395	1	R21	AA021225	2013	AA	Alcohol E	DESCR	2	\$SPECI	\$217,043	CALLACI, JOHN J.				
11	7981820	2	R01	A0531934	2009	AI	CDB+ T C	DESCR	2	\$SPECI	\$502,469	RIDDELL, STANLEY R.	JENSEN, MICHAEL C			
12	7731198	1	R01	CA136581	2009	CA	targeted	DESCR	2	\$SPECI	\$581,563	RIDDELL, STANLEY R. [contact]				
13	8371982	2	R01	DK079713	2012	DK	Project AI	DESCR	2	\$pecif	\$340,878	ARRIOLA, KIMBERLY RUTH JACOB				
14	7735633	2	R01	A0520794	2009	AI	Non-Hum	DESCR	2	\$pecif	\$400,000	KEARNS-JONKER, MARY K.				
15	8591825	1	R41	OD018402	2013	OD	CDX-301 i	DESCR	2	\$SPECI	\$100,000	MARSH, HENRY [contact]	YU, JIANNHA			
16	7785204	1	R01	N045109	2009	NS	A Ration	DESCR	A	Ration	\$338,726	BELLMKONDA, RAVI V.				
17	7741800	1	R01	DK083411	2009	DK	STEM CELL	DESCR	A	\$pec	\$376,800	LIN, FANGMING				
18	7811914	3	R01	DE014190	2009	DE	Injectable	DESCR	A	\$PECI	\$416,624	XU, HUAKUN				
19	7881433	1	R01	DK081118	2009	DK	A Rando	DESCR	A	\$pec	\$546,293	SIMMONS, LAURA A.				
20	8039487	1	R01	CA140243	2011	CA	HIC Deri	DESCR	A	\$pec	\$351,713	PAH, PINGPING				
21	7787991	1	R21	NR011192	2009	NR	Living Do	DESCR	A	\$PECI	\$244,000	TAYLOR, LAURA				
22	7566297	1	R01	DK082430	2009	DK	Histone p	DESCR	A	\$pec	\$377,500	KIKYO, NOBUAKI				
23	8371909	2	R01	DE013349	2012	DE	Engineeri	DESCR	A	\$pec	\$428,932	MOONEY, DAVID J.				
24	7804148	2	R44	HL071389	2010	HL	A Self-Mo	DESCR	A	\$pec	\$493,102	VILCOMBERON, DAVID [contact]	BLEBEA, JOHN			
25	8291578	1	R01	GM098294	2012	GM	Histone si	DESCR	A	\$pec	\$288,800	KIKYO, NOBUAKI				
26	8300575	1	K23	DK092029	2012	DK	Gender ca	DESCR	A	1	\$pe	\$182,110	FORDE, KIMBERLY AUTAMIN			
27	8392366	1	R43	AI1024494	2012	AI	Neutrophi	DESCR	Acute g	\$284,820	ROTOLO, JIMMY					
28	8394126	1	R41	DK095639	2012	DK	Measurem	DESCR	All form	\$584,931	HEROLD, KEVIN C					
29	8115617	2	R01	HL069229	2011	HL	Strategie	DESCR	Allogene	\$584,742	VAN DEN BRINK, MARCEL R M					
30	8185831	1	K08	CA140689	2011	CA	Strategie	DESCR	Among	\$127,683	ZAKRZEWSKI, JOHANNES					
31	8801263	1	K01	DK097194	2013	DK	Role of m	DESCR	B	\$PECI	\$138,305	ABDULREDA, MOHATH				
32	7831716	2	R01	NS044409	2009	NS	Bioengine	DESCR	Bioengin	\$384,794	BELLMKONDA, RAVI V.					
33	8256167	1	F30	HL112505	2012	HL	Biologica	DESCR	Cardiov	\$47,232	ENG, GEORGE					
34	7810378	1	P01	A0848534	2010	AI	Transgen	DESCR	The overc	Chronic	\$302,893	GREP, ANNE E				
35	8801263	1	K01	DK097194	2013	DK	Role of m	DESCR	B	\$PECI	\$138,305	ABDULREDA, MOHATH				
36	7713114	1	R21	A0840194	2009	AI	IL-10 & FC	DESCR	Despit	\$258,000	LA ROSA, CORRINA [contact]	LMAYE, ABHJIT PRAKASH				
37	8348583	1	R01	AI1014064	2012	AI	IL-22 in Th	DESCR	Endoge	\$502,673	VAN DEN BRINK, MARCEL R M					
38	8480026	1	K99	CA176376	2013	CA	Investigat	DESCR	Endothe	\$178,740	DUDAKOV, JARROD					
39	8189907	1	R21	DK095887	2011	DK	Endothe	DESCR	Endothe	\$190,625	BRODSKY, SERGEY					
40	7455670	1	R01	A055914	2009	AR	Gene Tra	DESCR	Gene tr	\$703,840	LANE, ALFRED T [contact]	KHAVA				
41	8537042	2	R56	A0645543	2012	AI	HCTIndi	DESCR	Hemato	\$363,157	MOORE, BETHANY S.					
42	8225815	1	K01	CA143999	2011	CA	Biology o	DESCR	Hemato	\$130,694	AGUILA, JERELL ROLAND					
43	8815891	2	R56	A0720394	2012	AI	Lymphoic	DESCR	Hypoth	\$283,750	BROMBERG, JOHATHAN S					
44	8234332	1	R01	DK091626	2011	DK	A Novel B	DESCR	titel from	\$340,093	OSERHOLZER, JOSE					
45	8333130	1	R01	HL114684	2012	HL	MicroRNA	DESCR	MicroRN	\$542,796	XU, MEIFENG [contact]	ASHRAF				
46	8313173	2	R44	HL097521	2012	HL	Treatmen	DESCR	Organ tr	\$706,733	TOY-MANNING, PAMELA ANN					
47	8125722	1	R43	DK092078	2011	DK	Anti-CD4	DESCR	Organ tr	\$274,232	TOY-MANNING, PAMELA ANN					
48	8645106	2	R44	DK092078	2013	DK	Anti-CD4	DESCR	Organ tr	\$499,799	TOY-MANNING, PAMELA ANN					
49	8031880	1	R21	A0925014	2011	AI	ROCK inh	DESCR	Out resp	\$264,150	ITYENGAR, SUJATHA					

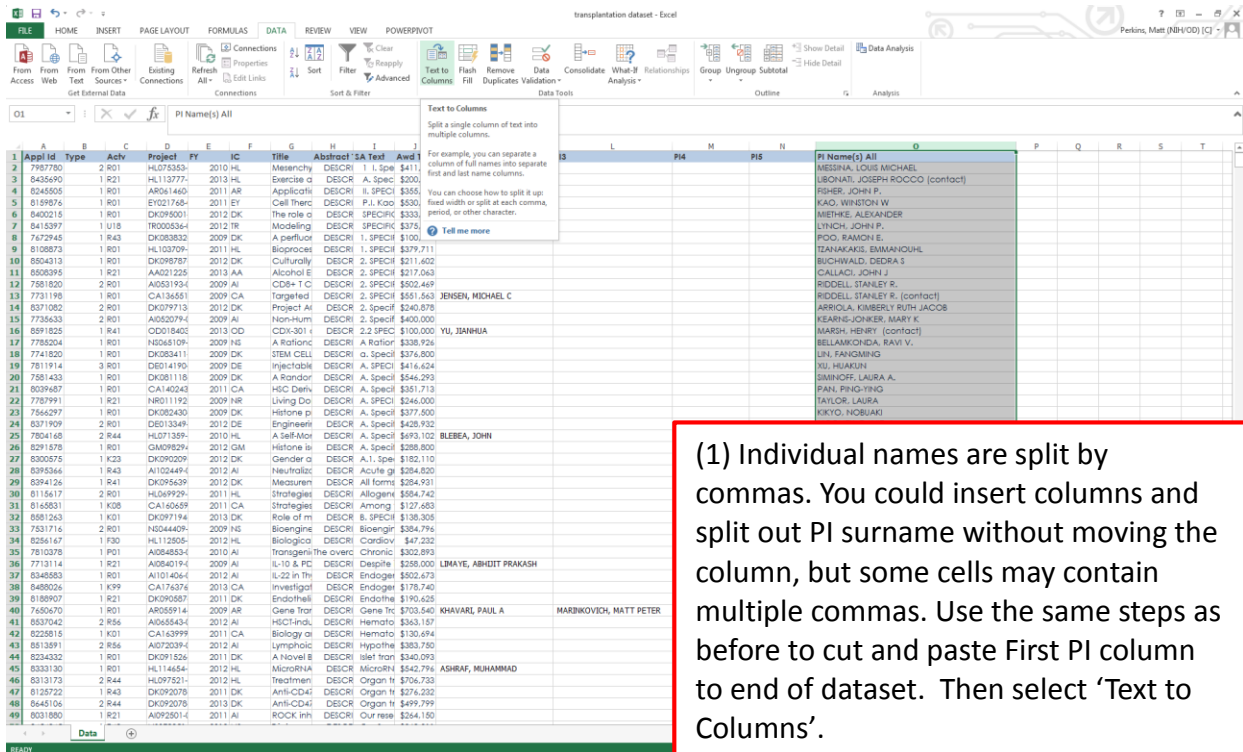
(8) Label the new columns, making sure that you label all columns containing data.

Example 2: Extracting surname from PI name

In this example the Transplantation dataset is used.

In the example above, the PI names were split into separate columns. But it may be that you are interested in just the PI surname. 'Text to Columns' can be used for this too.

An alternative would be to use LEFT and SEARCH to extract just the text up to the ',' which would leave the existing cell intact. (see help for more details)



The screenshot shows the 'Text to Columns' wizard in Microsoft Excel. The 'Text to Columns' tab is selected, and the 'Delimited' option is chosen. The 'PI Name(s) All' column is highlighted in the 'Text to Columns' list. A red box highlights the instruction: '(1) Individual names are split by commas. You could insert columns and split out PI surname without moving the column, but some cells may contain multiple commas. Use the same steps as before to cut and paste First PI column to end of dataset. Then select 'Text to Columns'.'

Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.
This is correct, choose Next, or choose the data type that best describes your data.

Original data type
PI Name(s) All

Choose the file type that best describes your data.

Delimited - Characters such as commas or tabs separate each field.
 Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

PI Name (s) All
MESSINA, LOUIS MICHAEL
LIBONATI, JOSEPH ROCCO (contact)
FISHER, JOHN P.
KAO, WINSTON W
METHHE, ALEXANDER
LYNCH, JOHN P.
POO, RAMON E.
TZANAKAKIS, EMMANOUIL
BUCHWALD, DEBRA S
CALLAGH, JOHN J
RIDDELL, STANLEY R.
RIDDELL, STANLEY R. (contact)
ARRICOLA, KIMBERLY RUTH JACOB
KEARNS-JONKER, MARY K
MARSH, HENRY (contact)
BELLAMKONDA, RAVI V.
LIN, FANCHANG
YU, HIAKUN
SHIMOFF, LAURA A.
PAN, PING-YING
TAYLOR, LAURA
KIKYO, NOBUAKI
MOONEY, DAVID J
VILKOMERSON, DAVID (contact)
KIKYO, NOBUAKI
FORDE, KIMBERLY AUTUMN
ROTOLO, JIMMY
HEROLD, KEVANI C
VAN DEN BRINK, MARCEL R M
ZAKREWSKI, JOHANNES
ABDULREDA, MICHAL H
BELLAMKONDA, RAVI V.
ENG, GEORGE
GRIEP, ANNE E
LA ROSA, CORINNA (contact)
VAN DEN BRINK, MARCEL R M
DUDAKOV, JARROD
BRODICKY, SERGEY
LANE, ALFRED I (contact)
MOORE, BETHANY B
AGUILA, JERRELL ROLAND
BROMBERG, JONATHAN S

Buttons: Cancel, Back, Next, Finish

(2) As before, ensure 'Delimited' is selected and click 'Next'.

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters
 Tab
 Comma
 Semicolon
 Space
 Other

Treat consecutive delimiters as one

Test qualifier: []

Data preview

PI Name (s) All
MESSINA, LOUIS MICHAEL
LIBONATI, JOSEPH ROCCO (contacts)
FISHER, JOHN P.
KAO, WINSTON W

Buttons: Cancel, Back, Next, Finish

(3) Then uncheck 'semicolon' and select 'Comma'. Then click 'Next'.

Convert Text to Columns Wizard - Step 3 of 3

This screen lets you select each column and set the Data Format.

Column data format

General
 Date
 Do not import column(s)

General converts numeric values to numbers, date values to dates, and all remaining values to text.

Destination: \$O\$1

Data preview

PI Name(s) All

MEISNA, LOUIS MICHAEL
 LUBONATI, JOSEPH ROCCO (contact)
 FISHER, JOHN P.
 KAO, WINSTON W
 MIETHE, ALEXANDER
 LYNCH, JOHN P.
 POO, RAMON E.
 TANAKAKI, EMMANOUEH
 BUCHWALD, DEBRA S
 CALLACI, JOHN J
 RIDDELL, STANLEY R.
 ARRIOLA, KIMBERLY RUTH JACOB
 KEARNS-JOHNKER, MARY K
 MARSH, HENRY (contact)
 BELLAMKONDA, RAVI V.
 LIN, FANGMING
 XU, JIAKUN
 SHIMNOFF, LAURA A.
 PAN, PING-YING
 TAYLOR, LAURA
 KIKYO, NOBUAKI
 MOCHNEY, DAVID J
 VILKOMERSON, DAVID (contact)
 KIKYO, NOBUAKI
 FORDE, KIMBERLY AUTUMN
 ROTOLO, JIMMY
 HEROLD, KEVIN C
 VAN DEN BRINK, MARCEL R M
 ZAKREWSKI, JOHANNES
 ABDULREDA, MIDHAT H
 BELLAMKONDA, RAVI V.
 ENG, GEORGE
 GRIEF, ANNE E
 LA ROSA, CORBINA (contact)
 VAN DEN BRINK, MARCEL R M
 DUDAKOV, JARROD
 BRODZKY, SERGEY
 LANGE, ALFRED T (contact)
 MOORE, BETHANY B
 AGUILA, JERELL ROLAND
 BROCKBERG, JOHANNAS J

(4) Again leave 'General' selected in the format options and click on 'Finish'.

transplantation dataset - Excel

PI Name(s) All

LOUIS MICHAEL
 JOSEPH ROCCO (contact)
 JOHN P.
 WINSTON W
 ALEXANDER
 JOHN P.
 RAMON E.
 EMMANOUEH
 DEBRA S
 JOHN J
 STANLEY R.
 KIMBERLY RUTH JACOB
 MARY K
 HENRY (contact)
 RAVI V.
 FANGMING
 LIN
 JIAKUN
 LAURA A.
 PING-YING
 LAURA
 NOBUAKI
 DAVID J
 (contact)
 NOBUAKI
 KIMBERLY AUTUMN
 JIMMY
 KEVIN C
 MARCEL R M
 JOHANNES
 MIDHAT H
 RAVI V.
 GEORGE
 ANNE E
 CORBINA (contact)
 MARCEL R M
 JARROD
 SERGEY
 ALFRED T (contact)

(5) Surname will remain in the original column and other names (and values) in the following columns.

The screenshot shows an Excel spreadsheet with a data table. The columns are labeled A through S. The data includes columns for 'AppId', 'Type', 'Actv', 'Project', 'PI Name(s) All', 'FY IC', 'Title', 'Abstract Text (oSA Text)', 'Awd Tot \$', 'IC (from PN)', and 'App Type'. A red box highlights a row where the name 'LAURENCIN' is split across columns M through S, with commas in the text.

(6) Check the results. In this case, there are PIs that have additional commas in their names resulting in text appearing in additional columns. If the interest is just in surnames this could be ignored, or the text could be concatenated.

More help

Click on the '?' in the top left (as below) and search for 'Text to Columns'.

The screenshot shows the 'Data' tab in the Excel ribbon. The 'Text to Columns' button is highlighted with a red box. The spreadsheet below shows a few rows of data with columns for 'AppId', 'Type', 'Actv', 'Project', 'PI Name(s) All', 'FY IC', 'Title', 'Abstract Text (oSA Text)', 'Awd Tot \$', 'IC (from PN)', and 'App Type'.