

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

Contents

Find a place for the woodchipper: Make a destination folder for csv files	1
Put the Branches into the Wheelbarrow: Covert pdfs to single excel file	2
Trim the Branches: Get data ready for Woodchipping	5
Move the rows	5
Add new columns (if needed)	6
Assemble the Woodchipper: Make the PowerQuery	8
Link the folder with csv files	8
Tell Excel What you want the PowerQuery to do	10
Save the Query and Run the Woodchipper	11
After the first time: Add new Files to the Woodchipper	12

Intro

A woodchipper is a semi-automated process to move data from filled out pdf forms into excel. This is useful for when you need to repeat the steps when new forms come in. Instead of copy and pasting each field from the pdf, you will export the data into excel, do some initial clean-up, run the woodchipper, and check for errors. The Woodchipper query is non-destructive. That means if you want to delete columns or filter in the query editor, the original data doesn't change.

Linking the Woodchipper to a folder means as new forms come in, you can add them to the query without having to touch the data you have already processed. New csv files have to have the exact same column names in the same order for the Woodchipper to work. Case matters - so if you change or add a column name as part of the trimming, you have to do that to each new csv file.

Find a place for the woodchipper: Make a destination folder for csv files

1. I like to title it something like "csv for woodchipper goes here"

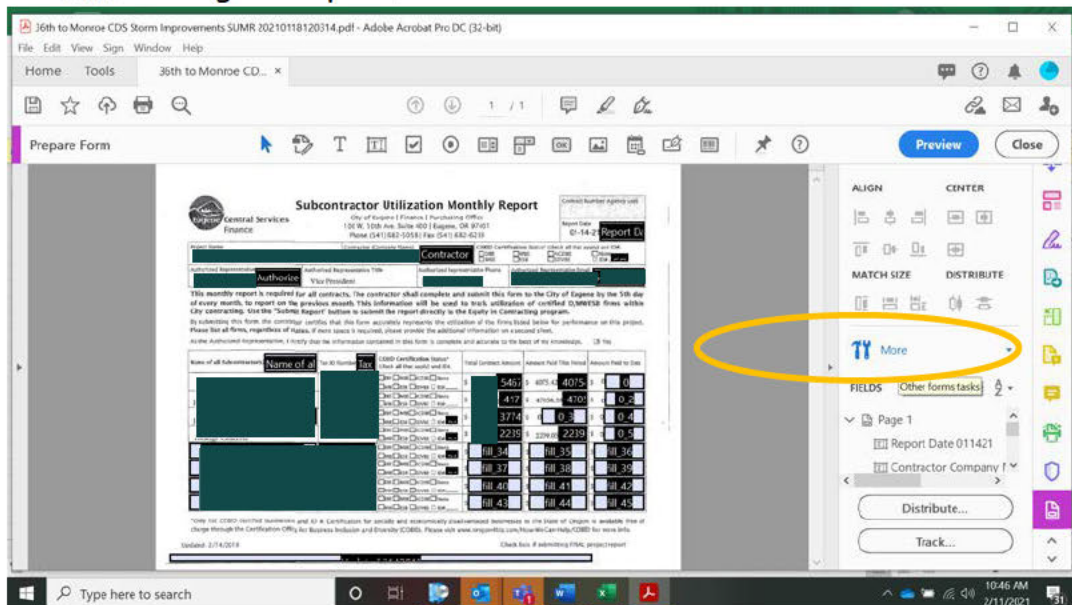
HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

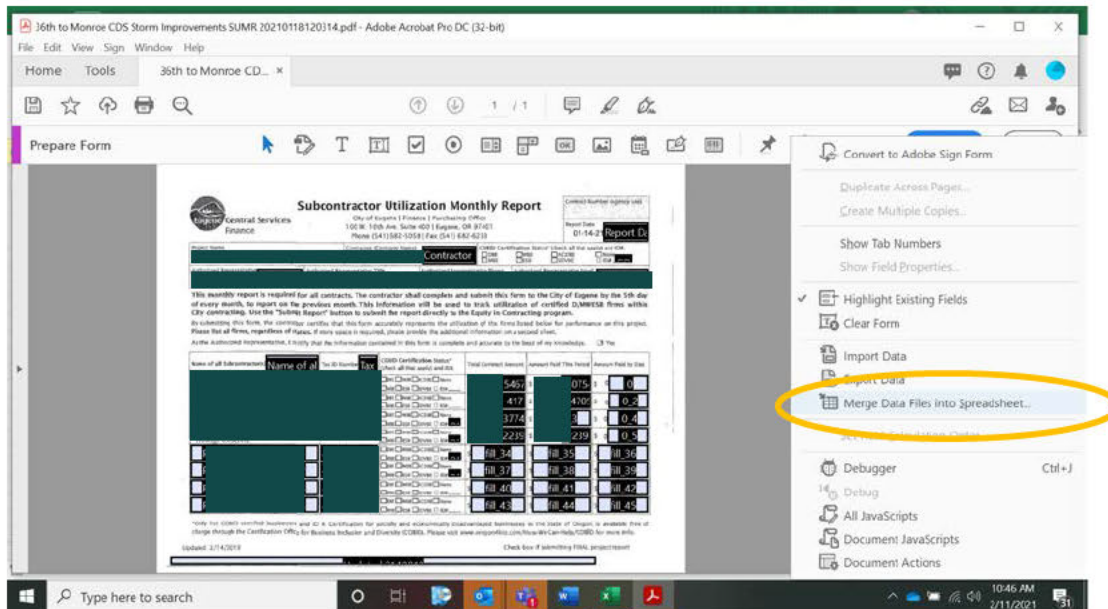
Equity in Contracting

May 2021

4. Click More to get dropdown menu



5. Select "Merge data into spreadsheet"



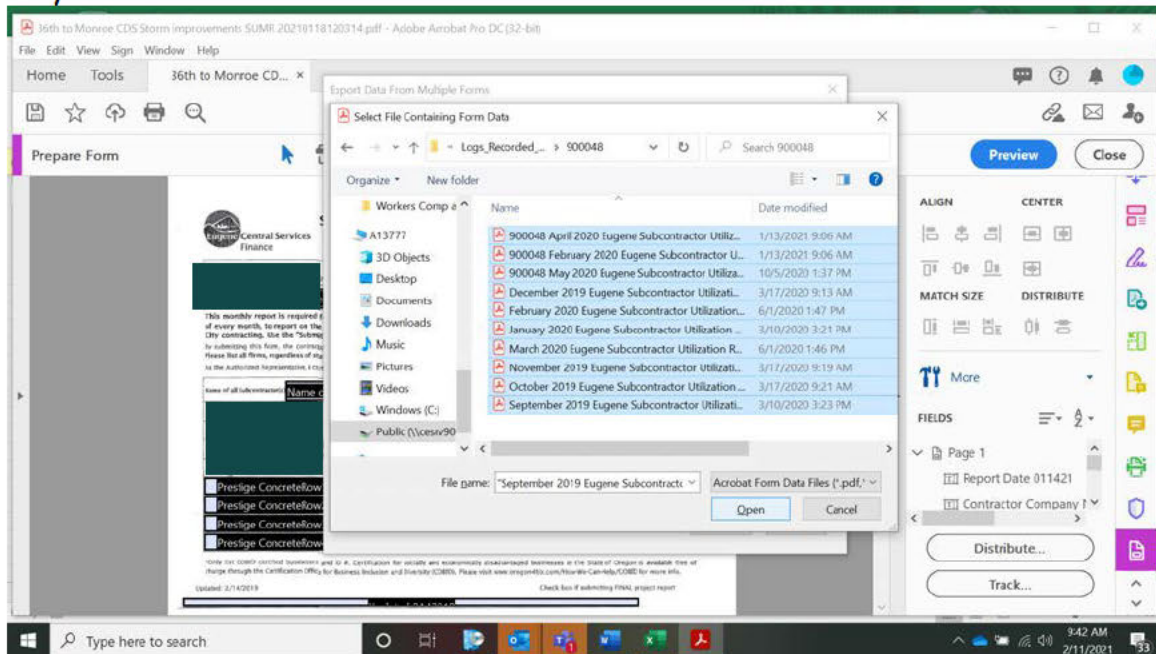
- a. In the pop up
 - i. Click add files
 - ii. Select the new files to be combined
 - iii. Click Open

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

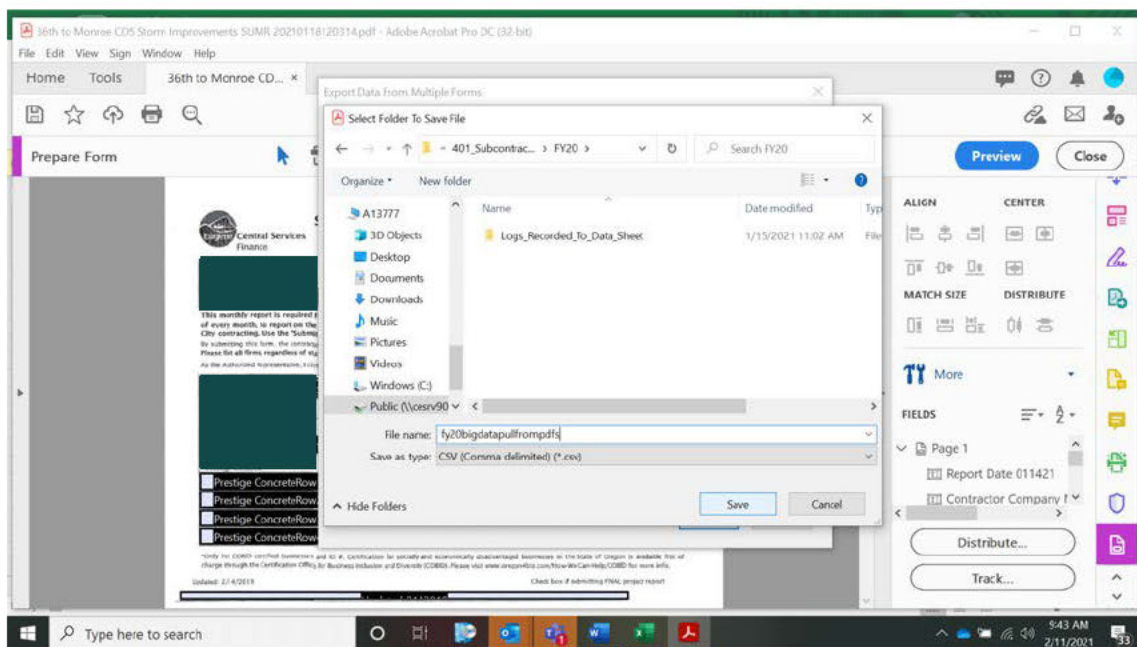
May 2021



b. In next pop-up

i. Name csv file

ii. Click Save



HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

Trim the Branches: Get data ready for Processing

Move the rows

1. Highlight the top row in one color
2. Highlight the last row in a second color
3. Select the header information
 - a. Copy and Paste as new rows - match the top line and bottom-line colors
 - b. Make 8 sets total
4. Select the Subcontractor info
 - a. Select Name of all Subcontractors Row2 to Amount Paid to Date2
 - b. Copy Ctrl X
 - c. Paste Cells in next open spot Ctrl V
 - d. Select the next set of subcontractor info and repeat

Contract Number (Agency use)	FISCAL_YEAR	Report Date	Project Name
900095 Eugene Subcontractor Utilization Report - Wildish Construction Co. - May 2020.pdf	900095 FY20	May-20	Courthouse District Vegetated Wil
COE Subcontractor Utilization Report JULY 2019.pdf	FY20	7/31/2019	Sladden Park Renovation, Cont 2G
COE Subcontractor Utilization Report JUNE 2019.pdf	FY20	6/30/2019	Sladden Park Renovation, Cont 2G
December 2019 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Dec-19	Fair Oaks Dr., Bedford Way an Wil
Facilities Use Only Subcontractor Utilization Report.pdf	FY20	9/7/2019	LCRS Holt Soreng Theater Me
February 2020 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Feb-20	Fair Oaks Dr., Bedford Way an Wil
ILR Subcontractor Utilization .pdf	FY20		
January 2020 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Jan-20	Fair Oaks Dr., Bedford Way an Wil
March 2020 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Mar-20	Fair Oaks Dr., Bedford Way an Wil
November 2019 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Nov-19	Fair Oaks Dr., Bedford Way an Wil
October 2019 Eugene Subcontractor Utilization Report.pdf	900048 FY20	Oct-19	Fair Oaks Dr., Bedford Way an Wil
September 2019 Eugene Subcontractor Utilization Report.pdf	900048 FY20	9/25/2019	Fair Oaks Dr., Bedford Way an Wil
SubUtilization (002) 2019-00245 Jan 2020.pdf	2019-00245	1/17/2020	Delta Ponds kpf
SubUtilization (002) babb 11 22.pdf	2019-00245	11/19/2019	Delta Ponds kpf
SubUtilization (002) babb 12 19.pdf	2019-00245	12/5/2019	Delta Ponds kpf

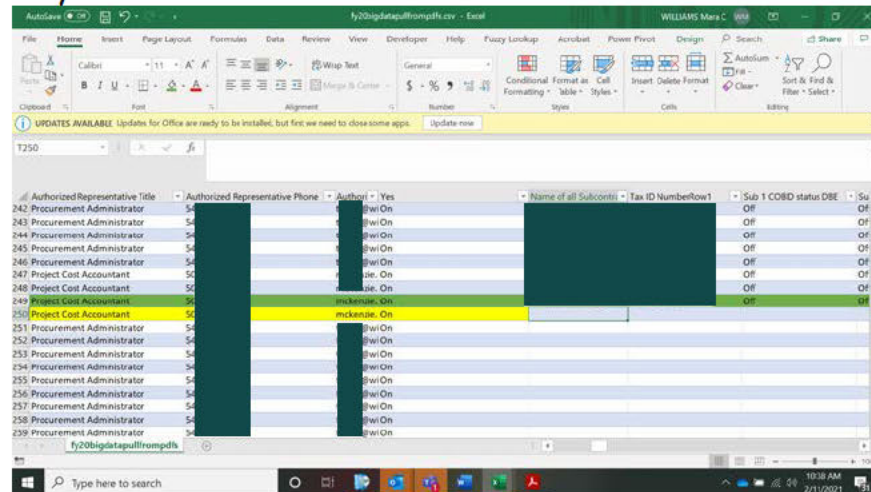
Green highlight is
bottom row

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

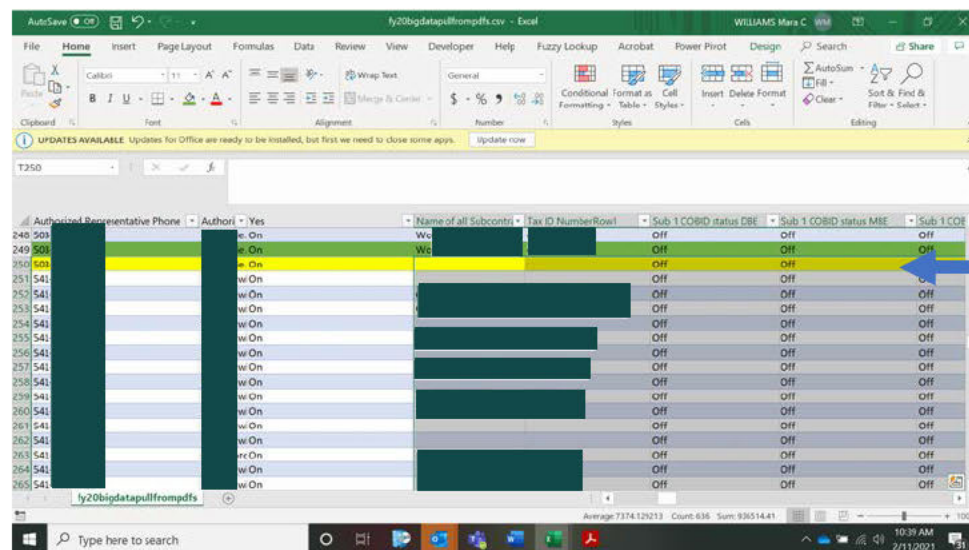
Equity in Contracting

May 2021



Excel spreadsheet showing a table with columns: Authorized Representative Title, Authorized Representative Phone, Author, Yes, Name of all Subcontr, Tax ID Number, Sub 1 COBID status DBE, and Sub 1 COBID status MBE. Row 249 is highlighted in yellow.

Yellow highlight is top row.



Excel spreadsheet showing a table with columns: Authorized Representative Phone, Author, Yes, Name of all Subcontr, Tax ID Number, Sub 1 COBID status DBE, Sub 1 COBID status MBE, and Sub 1 COBID status WBE. Row 249 is highlighted in yellow.

Match yellow highlight of top row to make sure the sections line up

Add new columns and change formatting (if needed)

1. Name first column
 - a. "File name"
2. Add fiscal year column
 - a. Insert column to left of "Contract Number (Agency use)"
 - b. Name it "FISCAL_YEAR"
 - c. Type in FY21
 - d. Double click tiny blue square in right bottom of cell
 - e. This gives you a series FY21, FY22
 - f. Click the tiny box at the bottom of the table

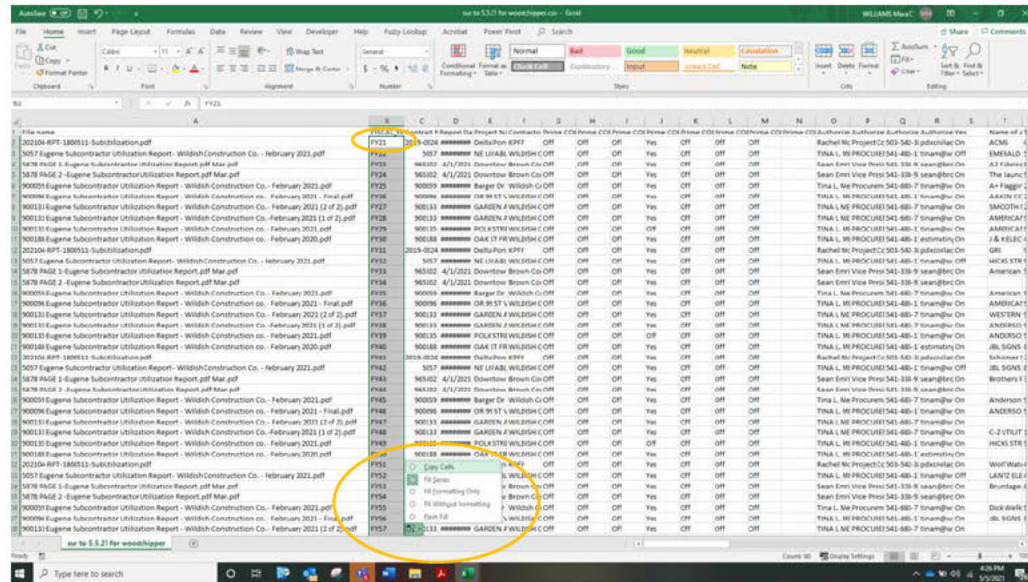
HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

g. Chose Copy Cells



3. Change date column formatting

- Highlight date column
- Go to Home tab on ribbon
- Switch format to "short date"

4. Remove bad rows

- If there are whole rows you want to leave out, you can delete them manually or wait and do it in the Woodchipper Query
 - If the bad row is only going to happen once (like a revised report), remove it here
 - If it's something you will want to do every time, like removing blank rows, do that later in the Woodchipper Query

5. Final check on column names:

- For the Woodchipper to treat every csv file the same the column names need to match (it's case sensitive) and in the same order For the Woodchipper to treat every csv file the same the column names need to be the same and in the same order
 - File name
 - FISCAL_YEAR
 - Contract Number (Agency use)
 - Report Date Project Name

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

- e. Contractor Company Name
- f. Prime COBID status DBE
- g. Prime COBID status WBE ...

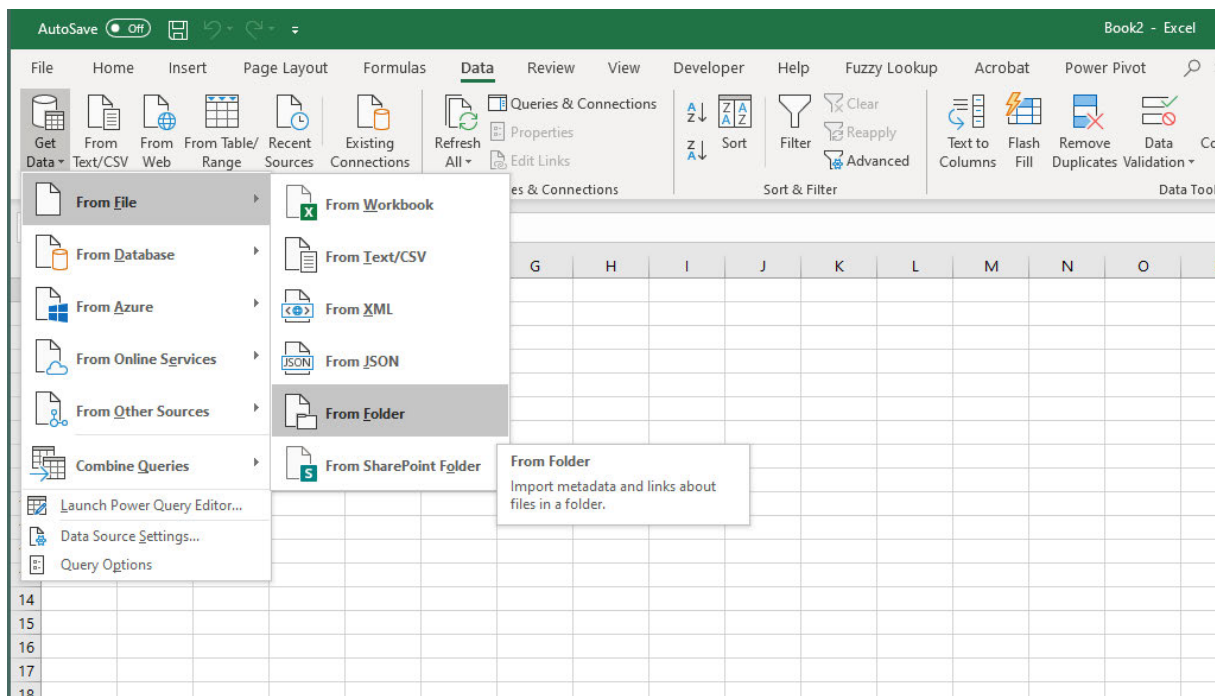
Save in Destination folder

1. Make sure you are in the right folder, "i.e. csv for woodchipper goes here"
2. File processed pdfs where you want to keep them

Assemble the Woodchipper: Make the PowerQuery

Link the folder with csv files

1. Open new excel file
2. Connect folder with csv files
 - a. Click Data Ribbon
 - a. Select Get Data Icon
 - b. Choose From File
 - c. Choose From Folder



3. In Pop-up
 - a. Select path to folder location

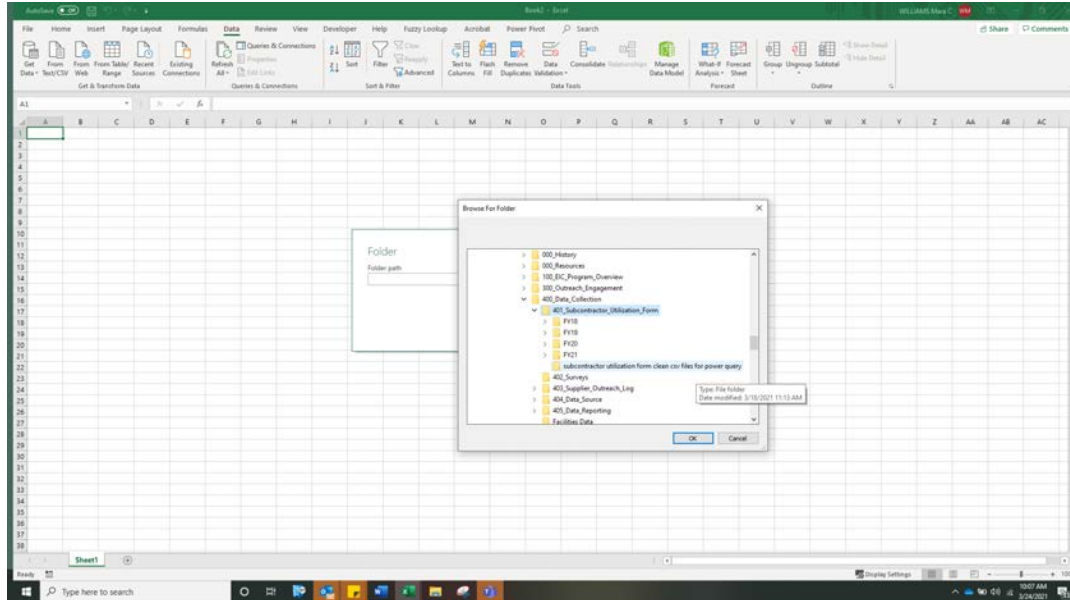
HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

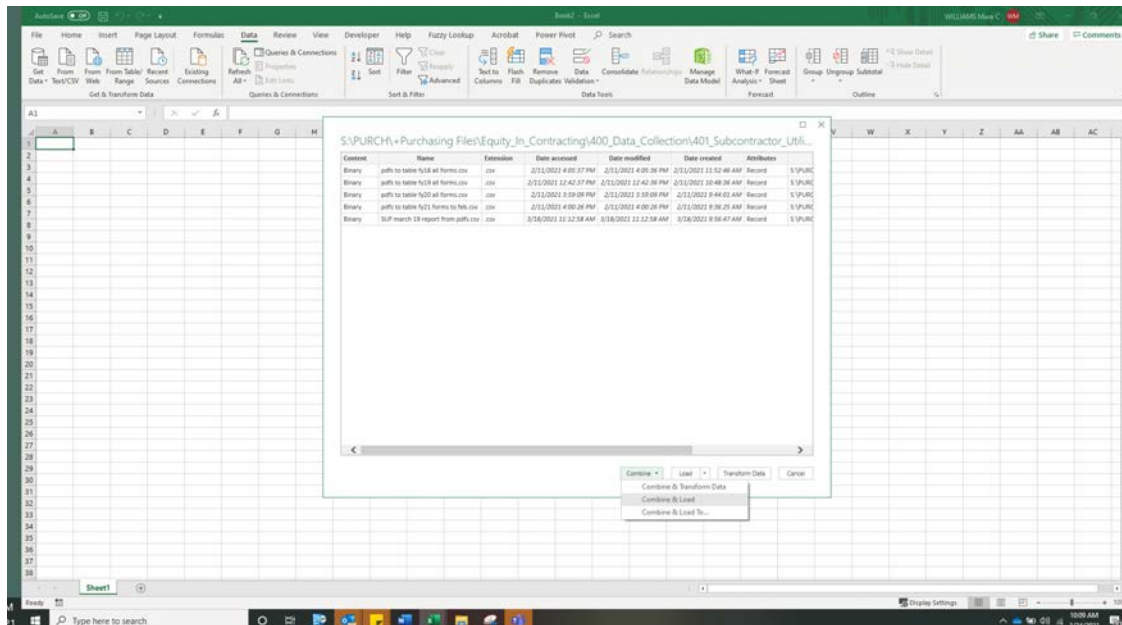
May 2021

- b. Select folder "csv files go here"
- c. Click okay



4. In next pop-up

- a. Click the down arrow next to Combine
- b. Select "Combine and Transform"



HOW TO BUILD A WOODCHIPPER IN EXCEL

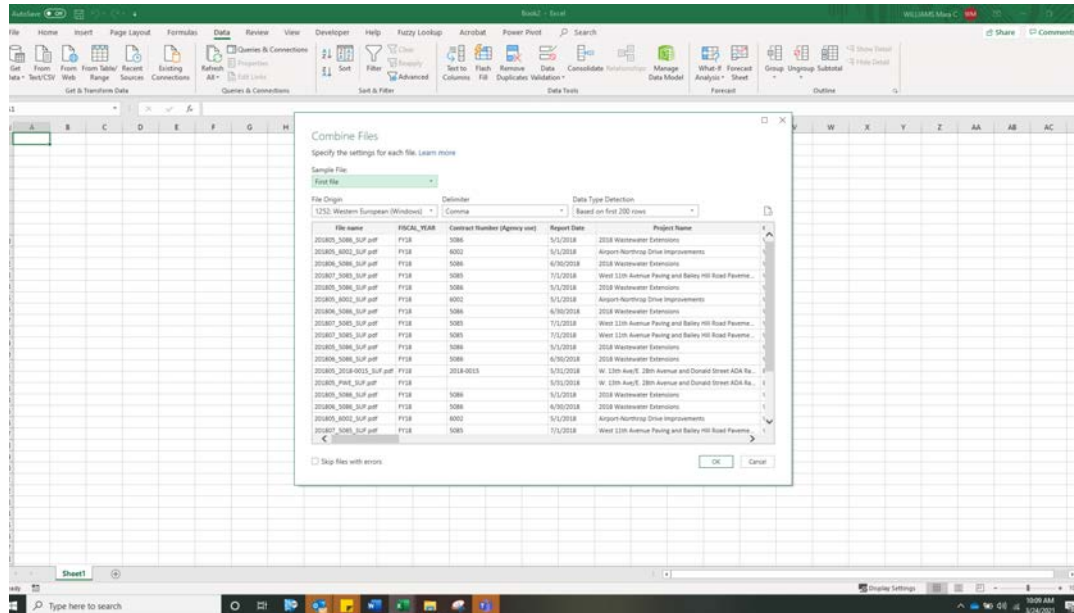
Mara Williams

Equity in Contracting

May 2021

5. In next pop up "Combine Files"

- Check to see if the data looks like what you expect
- Click ok



Set the dials: Tell Excel What You Want the PowerQuery to Do

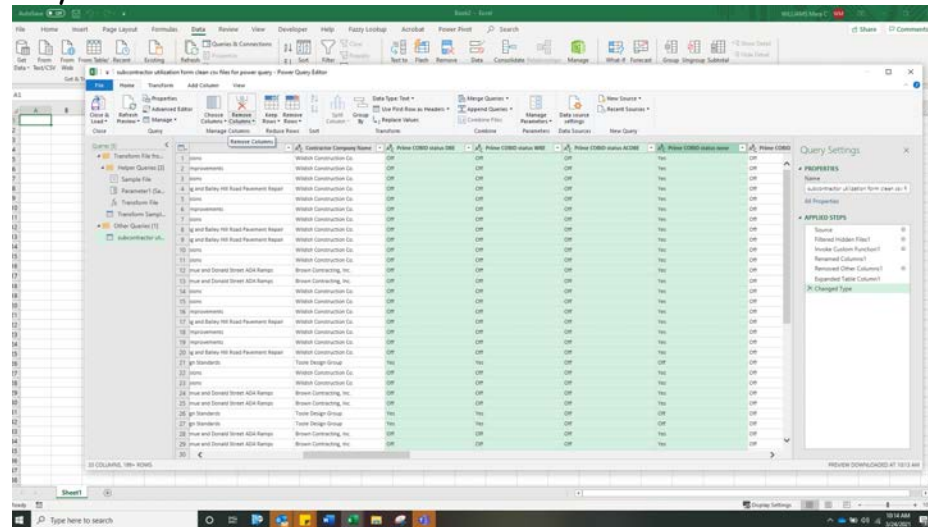
- Make sure top row has become column names
- Trim blank rows
 - Blank rows: right click on a column you know will not have any missing data (like Company Name)
 - Scroll through the choices and uncheck blank
- Unneeded columns: remove columns that you don't intend to use
 - Remove a single column
 - Select column by clicking top row
 - Hold shift to select multiple rows
 - Click Remove Columns Icon in Home Tab

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021



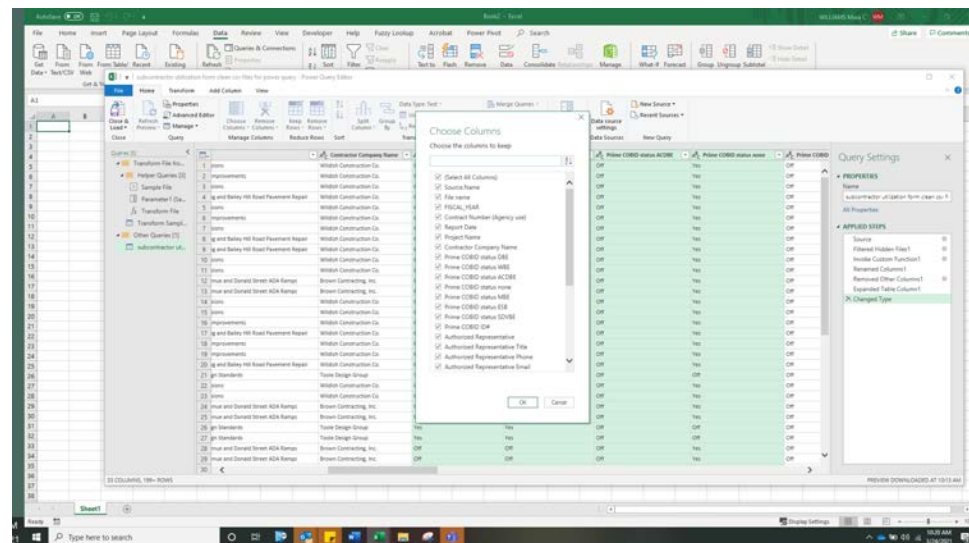
b. Remove multiple columns at once

i. In Home Tab click arrow drop down on Choose Columns

ii. Select "choose columns"

1. In pop-up check boxes for columns you want to keep

2. Click ok



4. Add an Index column

5. Rename your columns if you like

6. Add and transform columns as needed

Save the Query and Run the Woodchipper

a. If you want to see the results in a worksheet and work with it now:

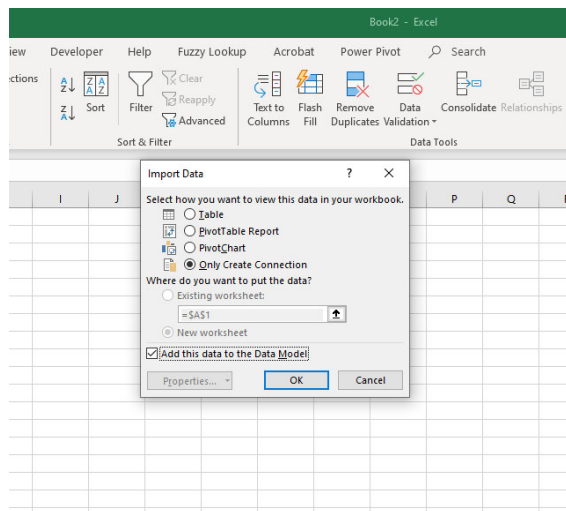
HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

- i. When you have finished making changes Click "Close and Load"
- b. If you are going to combine this with another query or relate it to other tables:
 - i. Click down arrow on "Close and Load"
 - ii. Select "close and Load to"
 - iii. In Pop-up "import Data"
 1. Select Only Create Connection
 2. Check "Add this data to Data Model"
 3. Click OK



Connection only is way friendlier on computer memory, so if you aren't going to use all the data in the query as it is, choose this option

After the first time: Add new Files to the Woodchipper

1. Repeat the steps to extract data from Adobe files and get data ready for woodchipper
2. Make sure the column names match your previous trimmed files
3. Save the file in the destination folder
4. Open the Woodchipper Workbook
5. Click Data Ribbon
 - a. Click Refresh All
 - b. You can tell it works if you look at the number of records or if you

HOW TO BUILD A WOODCHIPPER IN EXCEL

Mara Williams

Equity in Contracting

May 2021

Troubleshoot

The little yellow triangle: If you open your PowerQuery and see a yellow triangle - something has gone wrong.

- Open the PowerQuery Editor
- Click on the Applied Steps box on the right-hand side
- Either click go to error (which will take you to the right spot) or click each step starting from the top. The query will show what the data looks like as each step is applied. The first time you get no data/ the error screen is the step that is broken

Common problems that cause triangle error:

- The folder the PowerQuery is pulling from has been moved or restricted
- If your steps involve Merging two queries
 - i. the connection to the other query might be bad (i.e. that PowerQuery is broken)
 - ii. it might be that you changed column names or added/subtracted columns in the query you have merged into this one

Messy data in the forms themselves:

- Because Case matters in PowerQuery, different versions of the same Company Name like "123 Builders Co." and "123 Builders CO" and "123 builders co" (etc) will be treated like separate companies. I know they are the same company, but I need to tell excel that. There are multiple ways to solve this. Basically, you need to get the company names into a standard form.
 - Quick fix (but not for everything): You can highlight the column and change the case to All Caps. This will make both "123 Builders" and "123 builders" into "123 BUILDERS"
 - Better Fix: Make a name cleaning query.