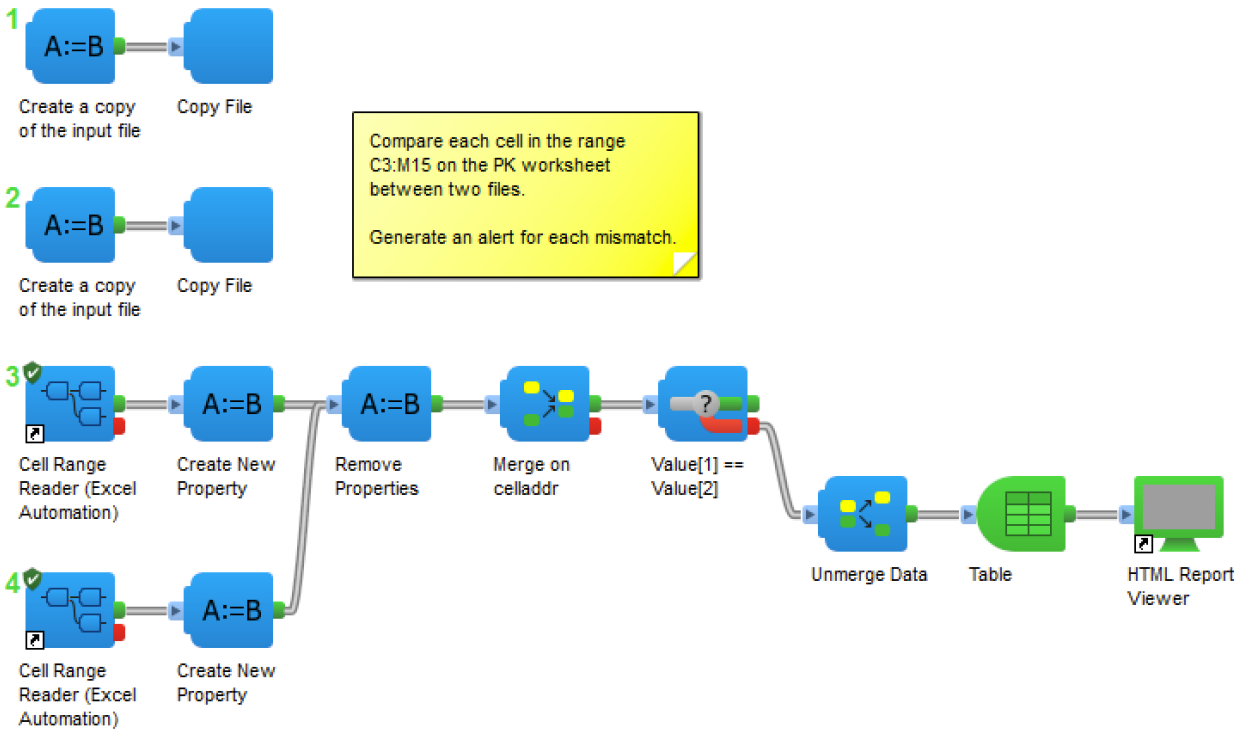
09 Reconcile Data Between Two Files

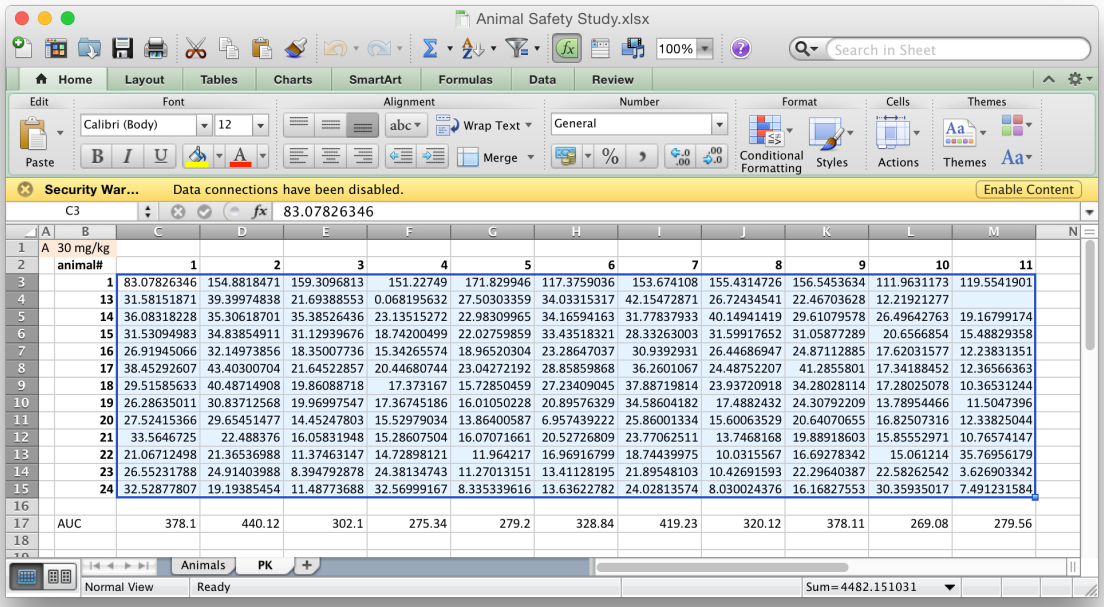
Purpose

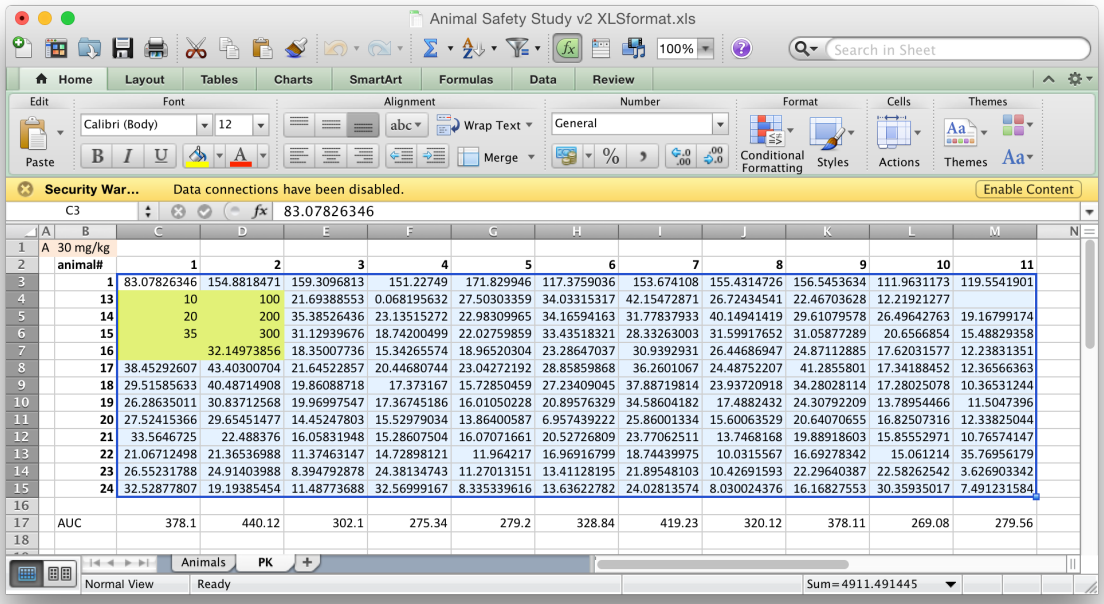
Automatically match values for each cell in the same cell range in two different Excel files, and identify discrepancies. Generate an alert for each mismatch.



Workflow

In this example we compare each cell in the range C3:M15 on the **PK** worksheet between two files: `data/Animal Safety Study.xlsx` and `data/Animal Safety Study v2 XLSformat.xls`.





Before reading any data we make a copy of both input files in the job directory. The original input files are shared by all users of the exercise: creating a copy helps ensure the original is not accidentally modified if you later decide to use a writer component on the file.

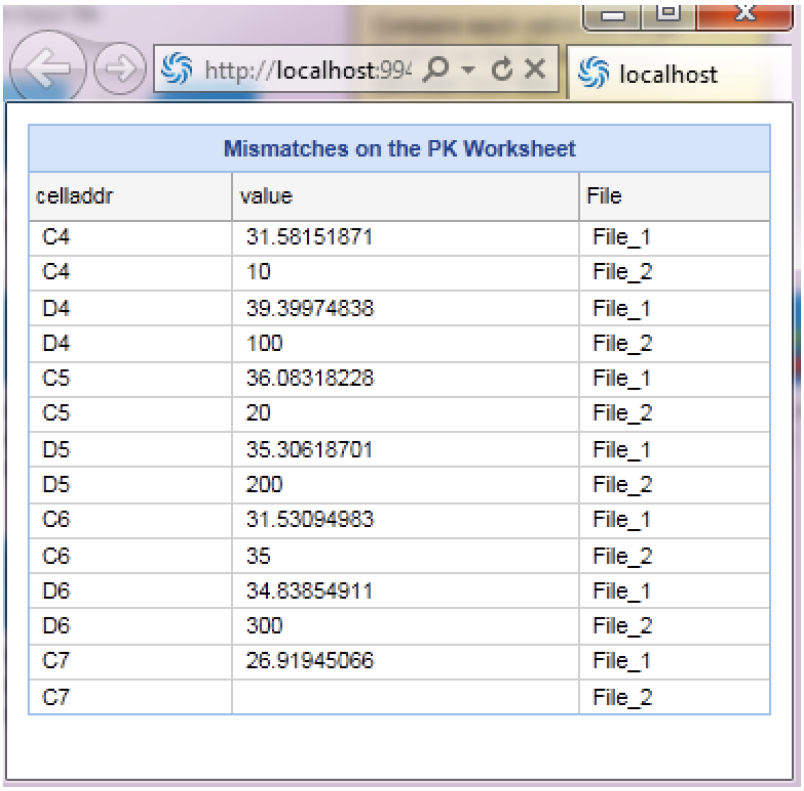
Using the **Cell Range Reader** component, we read from our copy of each file and extract the values from the C3:M15 range on the **PK** worksheet. We send them into Pipeline Pilot as data records with the property “value”.

We merge cell data on cell address (e.g. C8, E10, etc.) so that each cell pair becomes an array of values - as we are reading from two files.

Using a PilotScript filter component we compare the values in each pair and if they differ or one is missing, generate an alert in a web browser window.

Results

Here is the result of the protocol.



Mismatches on the PK Worksheet		
celladdr	value	File
C4	31.58151871	File_1
C4	10	File_2
D4	39.39974838	File_1
D4	100	File_2
C5	36.08318228	File_1
C5	20	File_2
D5	35.30618701	File_1
D5	200	File_2
C6	31.53094983	File_1
C6	35	File_2
D6	34.83854911	File_1
D6	300	File_2
C7	26.91945066	File_1
C7		File_2