

Handout

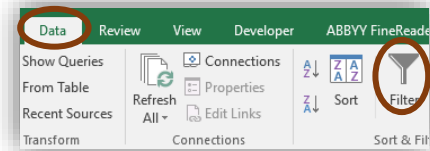
Data Remediation at Scale

How to Clean Up Your Metadata Quickly and Effectively Using Excel

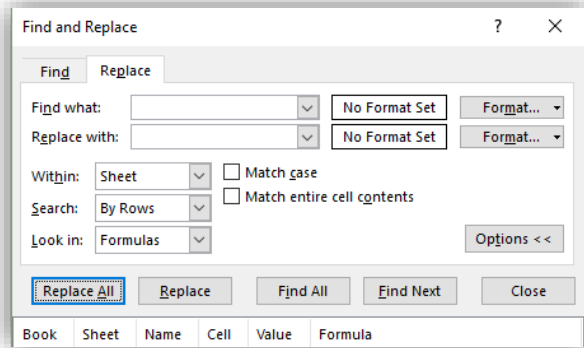
Marina Georgieva, Visiting Digital Collections Librarian

ADVANCED FEATURES

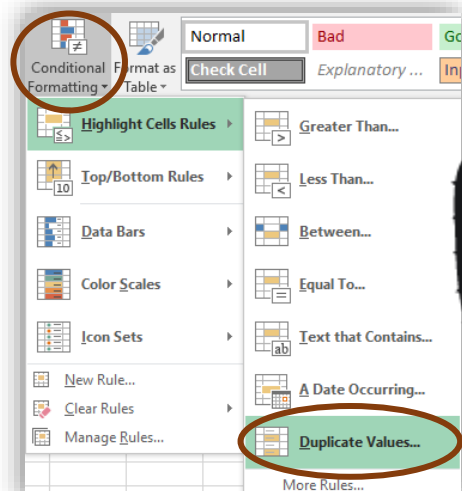
Data filter - sorts and groups data for manipulation



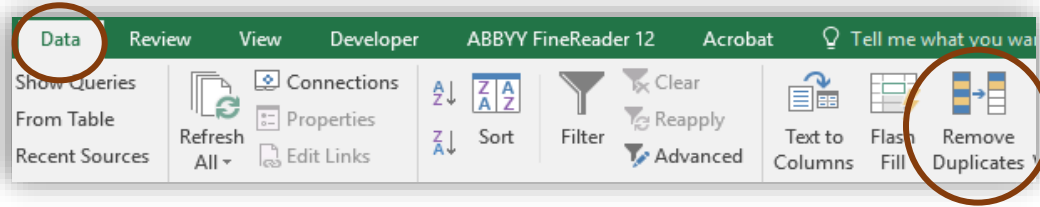
Find and replace - finds and replaces data by applying filters and setting ranges



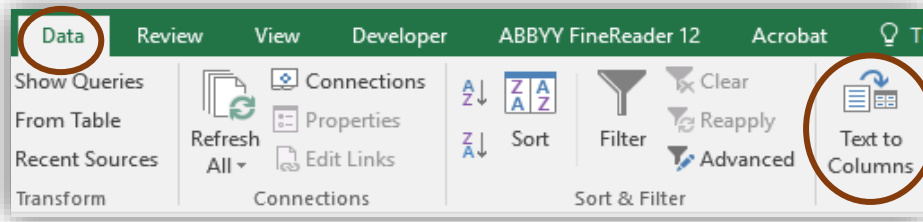
Duplicate values (look up) - identifies duplicates in specific fields or whole table



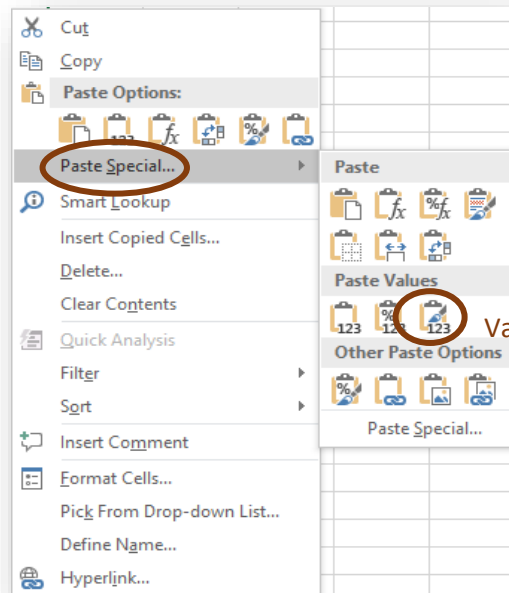
Duplicate values (remove) - identifies and removes duplicates in columns (*great for cleaning Subject field*)



Text to column - breaks delimited text strings to multiple columns (*great for cleaning Subject field*)



Date normalization - pastes dates in the preferred format without turning them into random numbers (*great for cleaning Date field*)



Values and source formatting



ADVANCED FUNCTIONS

TRIM | returns a text value with removed leading and trailing spaces

IF | logical function; removes end punctuation if a condition is met

CHAR(10) | inserts line break (new line)

MID | **LEFT** | **RIGHT** | returns specific number of characters from a text string; breaks the string

CONCATENATE | merges values from specified range of columns and delimits by punctuation

TEXTJOIN | merges values from specified range of columns and delimits by punctuation

*** The difference**

Both will join values across multiple columns, but

- *Concatenate* requires manual selection of cells
- *Textjoin* automatically merges values from a specified range (more convenient and powerful for large data sets)



INDEX and MATCH | indexes a specified cell, matches the value from a pre-defined table and returns requested value

Date normalization | specifies the preferred date format, transforms and keeps any date in the preferred format

Add string to value | add text string to existing value in a cell (*great for description field or any field with non-normalized data*)

