Wikidata Batch Editing and Creation Using QuickStatements

QuickStatements: https://quickstatements.toolforge.org/ is a tool that you can use to batch add and edit Wikidata pages. Note that you need to have a Wikidata account that is over 4 days old and has over 50 edits.

Advantages:

a. Can setup statements using existing tools (Excel)
b. Ideal to use when you have a group with identical statements (such as faculty from same department)
c. Tool helps avoid repetitive clicking needed to add statements manually
d. Can keep track of history of batches.
e. Will not add duplicate statements or entities
f. Will show you any errors

Disdvantages:

a. Can take a little bit of manual work to format the statements if you do it in Excel.
b. Can reduce work a little via OpenRefine, by using it to reconcile and obtain the values (Q statements), but needs to be very specific and in the format that Wikidata expects. Can use more than one value to reconcile.
c. Tool can occasionally run slowly depending on what else is in the pipeline
d. Hard to tell where errors are

Batch adding statements to already existing entries

1. In order to be properly formulated, the data needs to be arranged with their codes with tabs between entity/property/value. This can be easily done by recording the data in Excel (which automatically adds the tab when you copy/paste columns). Here's an example of the addition of "official websites" to these academic departments:

Q99737497	P856	"https://www.appliedarts.txstate.edu/"
Q99737522	P856	"https://www.mccoy.txstate.edu/"
Q99737523	P856	"https://www.education.txstate.edu/"
Q99737525	P856	"https://www.finearts.txstate.edu/"
Q99737527	P856	"https://www.health.txstate.edu/"
Q99737529	P856	"https://www.liberalarts.txstate.edu/"
Q99737530	P856	"https://www.cose.txstate.edu/"

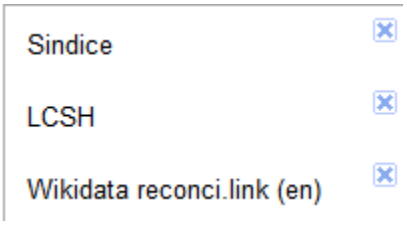
   a. As you can see, any type of text needs to have quotations around it.
   b. Before using the Excel spreadsheet, select the whole spreadsheet and format it as text.
   c. To add a source to the statement, use the property with a "S" instead of a "P" and formulate it with the following time format changing the date as appropriate: +2020-10-03T00:00:00Z/11. In the following, P854 is reference URL and P813 is retrieved. (See section titled "Adding Dates" for more information.)

P31	Q43229	S854	"https://www.hhp.txstate.edu/"	S813	+2020-10-03T00:00:00Z/11

2. Copy and paste the rows from Excel into QuickStatements: https://quickstatements.toolforge.org/ (You'll want to login to your user account). Also, here is the Help page: https://www.wikidata.org/wiki/Help:QuickStatements
3. Click "New Batch"
4. Paste your rows into the box.
5. Click "Import V1 commands"
6. The program will then evaluate your statements to see if there are any errors. If not, click "Run in background".
7. Depending on what is in the queue (I think), the program will either add your statements quickly or will take awhile. You can always click "Last batches" to see what activity there is.
8. Here's an excellent video about this process: https://youtu.be/Ql7gC91eWss

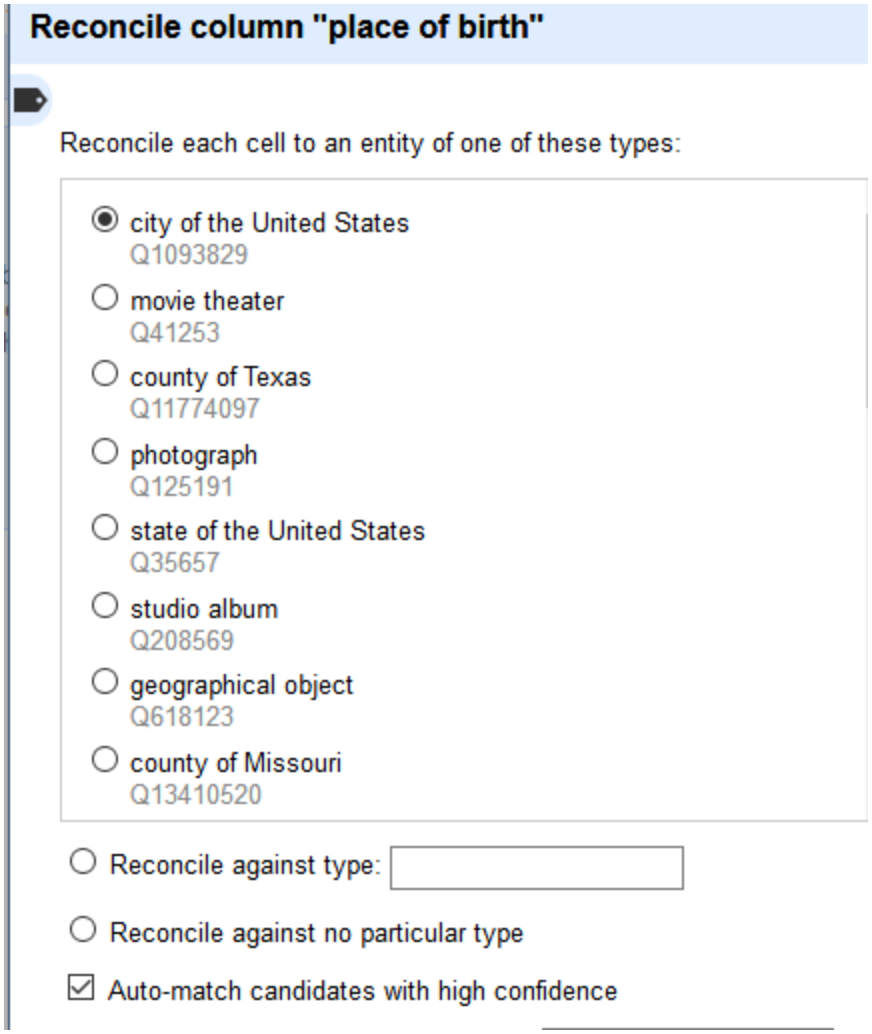
Reconciling data in OpenRefine to Find Q Numbers

1. You can always search Wikidata for Q numbers of properties (https://www.wikidata.org/wiki/Wikidata:List_of_properties) but that can be time-consuming also. A shortcut you can take is to reconcile the values through OpenRefine.
2. Install and open OpenRefine: https://openrefine.org/download.html (Might need to give it special permissions if you use a computer on the university network).
3. Import your data and create project (this example consists of places of birth).
    a. Enter N/A or something similar in blank cells.
4. Click the triangle at the column header, select Reconcile…Start Reconcile.
5. The default wikidata service no longer works. You can try it, but if nothing happens, you'll need to delete that service by clicking the X at the far right:

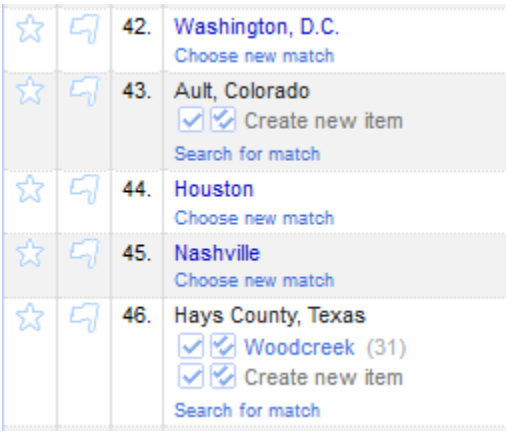


6. Click "Add Standard Service…" and add this new link (https://wikidata.reconci.link/en/api) .

7. The next screen will try to guess what entity it is. You can see below that it correctly guessed for our data:



8. Click “Start Reconciling”



9. Some of the cities were correctly reconciled (now in blue) but some were not (still in black), partially because these weren’t all cities (some counties/states). This data will need to be resolved (of course it was just a data dump as we were trying to figure it all out!)

10. For the ones that were correctly reconciled, click the column triangle again and “Reconcile…Add Entities Identifier Column.” The next screen will ask for the name of the colunn.
11. Click Export at the top right of the screen and whatever format you want.
12. Copy the cells into columns at the right of the column being examined.*

*We could just import the whole spreadsheet into OpenRefine and reconcile each column as needed, but I just tackled this one column as an example.

Batch Creating Entities

Very similar to batch editing in QuickStatements, except you add “Create” at beginning and “Last” in front of each statement:

CREATE						
LAST	Len	"Texas State University Health and Human Performance"				
LAST	Den	"academic department at Texas State University"				
LAST	P31	Q43229	S854	"https://w	S813	+2020-10-03T00:00:00Z/11
LAST	P31	Q2467461	S854	"https://w	S813	+2020-10-03T00:00:00Z/11
LAST	P361	Q1495387	S854	"https://w	S813	+2020-10-03T00:00:00Z/11
LAST	P361	Q9973752	S854	"https://w	S813	+2020-10-03T00:00:00Z/11
LAST	P101	Q2330929	S854	"https://w	S813	
LAST	P856	"https://www.hhp.txstate.edu/"				

Hyperlinks

All links must have quotation marks around it. But Excel has default settings to make all links into hyperlinks which does not fly in QuickStatements. See this helpful page that showed how to turn off the automatic hyperlinking in Excel: https://www.thespreadsheetguru.com/blog/prevent-auto-hyperlinking-in-excel

Adding dates

Exact dates are formatted +YYYY-MM-DDT00:00:00Z/PRECISION

Ex. For January 8, 1989 enter: **+1989-01-08T00:00:00Z/11**, where /11 designates the precision (day)

If only have a year format: **+1839-00-00T00:00:00Z/9**, where 9 indicates the precision (year)

The precision is: 0 - billion years, 1 - hundred million years, ..., 6 - millennium, 7 - century, 8 - decade, **9 - year** (default), 10 - month, 11 – day.
See this link for more information: https://www.wikidata.org/wiki/Help:QuickStatements