

Instructions to Use “acronyms_cleanup.py” to update acronyms:

Step 1. Run acronyms_list.py

```
[root@centos ~] -> python acronyms_list.py  
Enter Username for Mysql:ysui  
Enter Password for Mysql:  
2025 datasets have been checked
```

We will get an output file named “acronyms_list.csv”

Step 2: Upload the spell_out_list.csv (not required)

(From Local) > scp -P 23 <spell_out_list.csv> <username>@sferic:<directory for script>.

```
$ scp -P 23 ~/Desktop/result/spell_out_list.csv ysui@sferic:~/
```

Step 3: Run acronyms_cleanup.py

There will be 4 scenarios:

1. According to the spell_out_list.csv, the spell-out of the acronym in the dataset summary is correct. The dataset will be skipped without user's response:

```
*****Archive_ident = 106.427*****  
Over a three year period, scientists from the University of Wyoming and the United States Geological Survey (USGS) followed the movements of bears in both habitats and monitored their body temperature, muscle condition, blood chemistry, and metabolism  
*****  
'USGS: United States Geological Survey (USGS)' is correct, 106.427 is skipped
```

If the acronym does not need to be spelled out, the output will be like:

```
*****Archive_ident = 106.400*****  
The data are divided into two ZIP files consisting of yearly Excel files.',)  
*****  
'ZIP: ' is correct, 106.400 is skipped
```

2. According to “spell_out_list.csv”, the acronym in the dataset has record and can be spelled out automatically. The script will ask if you want to spell out automatically.
3. There is a spell-out of the acronym in the spell_out_list.csv, but the variation in the dataset summary has never been met:

For Example

In csv list, "DEM" has spell-out, but has no record of variations.

acronym	spell_out	variation
PAR	photosynthetically active radiation (PAR)	PAR (photosynthetically active radiation)
DGPS	Differential Global Positioning System (DGPS)	
BC	black carbon (BC)	BC (black carbon)
MSS		
14CH4		
EOS		
AMSR-E	Advanced Microwave Scanning Radiometer - EOS (AMSR-I AMSR-E (Advanced Microwave Scanning Radiometer - EOS)	
ACP	Arctic Coastal Plain (ACP)	
NetCDF		
DEM	Digital Elevation Model (DEM)	

When I run the program, I meet "digital elevation model (DEM)" which is different from the correct spell-out "Digital Elevation Model (DEM)" in the csv file. The program will ask you if you want to spell out DEM manually.

```
*****Archive_ident = 106.ARCSS301*****
This data set is a 100-m resolution digital elevation model (DEM) for the Brooks Range and the North S
lope of Alaska
*****
Do You Want to Spell Out 'DEM' manually[Y/N]:Y
Spell out of Acronym 'DEM': Digital Elevation Model
Replace Part: digital elevation model (DEM)
Do You Want to replace 'digital elevation model (DEM)' with 'Digital Elevation Model (DEM)' [Y/N]:Y
```

If the update is correct, the program will ask if you want to update the output file.

```
*****Archive_ident = 106.ARCSS301*****
This data set is a 100-m resolution Digital Elevation Model (DEM) for the Brooks Range and the North S
lope of Alaska
*****
Do You Want to Update Exprot File[Y/N]:Y
```

The variations of 'DEM' will be updated:

DEM	Digital Elevation Model (DEM)	digital elevation model (DEM)
-----	-------------------------------	-------------------------------

4. There is no record of the acronym in the spell_out_list.csv, the script will ask if you want to spell out manually and create a record in the spell out list

```
*****Archive_ident = 106.358*****
"This data set contains paleoceanographic data from core HLY03-05 GC taken during cruise HLY03 to Nare
s Strait in 2003, part of the Canadian Archipelago Throughflow Study (CATS)
*****
Do You Want to Spell Out 'CATS' manually[Y/N]:Y
Spell out of Acronym 'CATS': Canadian Archipelago Throughflow Studv
Replace Part: Canadian Archipelago Throughflow Study (CATS)
Do You Want to replace 'Canadian Archipelago Throughflow Study (CATS)' with 'Canadian Archipelago Thro
ughflow Study (CATS)' [Y/N]:Y
```

If the update is correct, the program will ask if you want to update the output file.

```
*****Archive_ident = 106.358*****
"This data set contains paleoceanographic data from core HLY03-05 GC taken during cruise HLY03 to Nare
s Strait in 2003, part of the Canadian Archipelago Throughflow Study (CATS)
*****
Do You Want to Update Exprot File[Y/N]:Y
```

The acronym and its spell-out will be inserted into the `spell_out_list`:

CATS Canadian Archipelago Throughflow Study (CATS)

If the acronyms has no spell out in the csv file, and when you first meet the acronym you answered “No” for manually clean up request, the acronym will be assumed to be “do not need to be spelled out” and it will be skipped in the following running.

Format of imported csv file:

acronym	spell_out	variation
PAR	photosynthetically active radiation (PAR)	PAR (photosynthetically active radiation)
DGPS	Differential Global Positioning System (DGPS)	
BC	black carbon (BC)	BC (black carbon)
MSS		
14CH4		
EOS		
AMSR-E	Advanced Microwave Scanning Radiometer - EOS (AMSR-E)	AMSR-E (Advanced Microwave Scanning Radiometer - EOS)
ACP	Arctic Coastal Plain (ACP)	
NetCDF		
DEM	Digital Elevation Model (DEM)	

1. The name must be “spell_out_list.csv”
 2. 3 columns: “acronym”, “spell_out”, “variation”
 3. if the spell_out column are left empty, means the acronyms does not need to be spelled out

Need to pay attention:

Some acronyms will not be covered by the script, such as two acronyms in the parentheses.

For example:

*****Archive_ident = 106.343*****
Paleoshorelines of Drained Thaw Lake Basins (DTLBs) on the Alaskan North Slope were mapped using the Interferometric Synthetic Aperture Radar Digital Elevation Model (InSAR DEM) and remote sensing imagery

Do You Want to Spell Out 'DEM' manually[Y/N]:N

For the cases like (InSAR DEM), the best way is to answer No here and go to Zinc to update the dataset description manually.

To download csv files: