# Problem Statement

You have been given a listing of anatomy terms and taxon stage ranges. The anatomy ontology developers would like to compare and contrast anatomy term usage across the different species. They have asked you to prepare the data so that they can review how the anatomy terms are used by each species.

ecao\_listing\_test20240222\_taxon\_all\_v1b.tab

# Task

### Using jupyter notebook load the data file

ecao\_listing\_test20240222\_taxon\_all\_v1b.tab

### Arrange the data using pivot or joins so that it can be viewed to compare how the anatomy terms are used by each species.

### Prepare a pull request containing your jupyter notebook and include the final display format of the data in a tab delimited file for the curators to review. Provide a description of the problems that you ran into while performing this task and how you approached solving them. Explain the various options you considered and how you identified the best approach to take.

## reference

<https://github.com/pellst/xbdevops_testcase/tree/main/data_wrangling/ecao_compare_in_taxon>

data file located here

<https://github.com/pellst/xbdevops_testcase/blob/main/data_wrangling/ecao_compare_in_taxon/data/ecao_listing_test20240222_taxon_all_v1b.tab>

The original query used to obtain the data, for information purposes only.

-- full list phylonode\_anatomyitem

with m1 as (

select t1.\*

, (select ai.name from anatomy\_item ai where ai.anatomy\_item\_id = t1.anatomy\_item\_id ) anatomy\_name

, ( select org.abbreviation from organism org where org.phylonode\_id = t1.phylonode\_id ) organism\_name

, (select st.name from stage st where st.stage\_id = t1.starts\_stage\_id ) starts\_at\_stage\_name

, (select st.name from stage st where st.stage\_id = t1.end\_stage\_id ) ends\_at\_stage\_name

from phylonode\_anatomyitem t1

--where t1.phylonode\_id = 560 -- S.pur -- org id 12 -- 7668 ncbitax ntx

where t1.phylonode\_id != 608 -- Perischoechinoidea -- not setup in organism table

)

select m1.anatomy\_item\_id, m1.anatomy\_name

, m1.phylonode\_id , m1.organism\_name

, m1.starts\_at\_stage\_name, m1.ends\_at\_stage\_name

--, m2.phylonode\_id pliv, m2.starts\_at\_stage\_name, m2.ends\_at\_stage\_name

, m1.phylonode\_anatomyitem\_id

from m1

order by m1.phylonode\_id, m1.anatomy\_item\_id

;