Homework Assignment 8

[Homework Assignment 8.docx](https://uncc.instructure.com/courses/192538/files/20593306?wrap=1)

This is a group assignment! Please keep the same partners from HW4 unless you have a good reason to change your group. When you submit this assignment, list the students participating in your group.

To make the grading simpler, I ask you to copy the database description and the Entity-Relationship diagram from HW4. However, you may edit/change them if necessary.

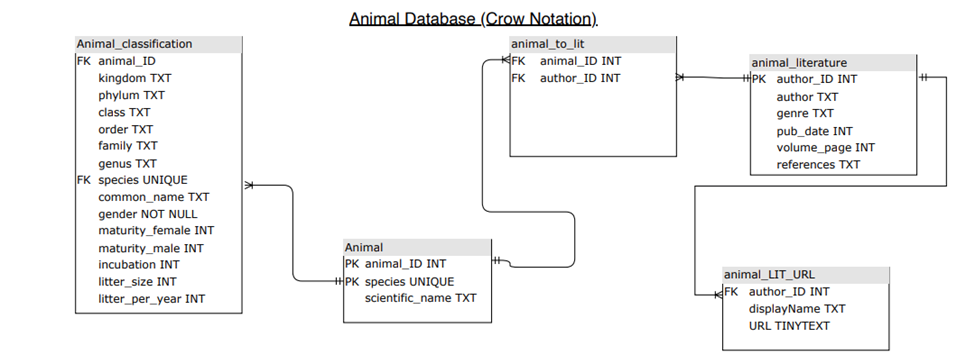
# Question 1 (10pts).

# Copy the database description from Question 1 of HW4. Edit/update if necessary.

Our original database would include the sequencing of a zoo with the related literature that is stored about each zoo animal. The sequencing of the animals that reside inside of the zoo allows the researchers access to valuable data that can be used by the likes of veterinarians and students. Having the ability and understanding of the inner workings of the zoo animals is key for many reasons including understanding future outbreaks (like we experienced with covid), understanding the genetic underpinnings of an organism’s traits and its interaction within a simulated ecosystem. This could be expanded upon in the future if another outbreak happens or what kinds of animals are susceptible to new arising illnesses and developing more robust/data about animals as an open source project. This database bridges the gap from the resource information to each zoo animal. Out of these combinations of resource information we can allow access from broad to more niche information about each animal that is available.

# Question 2 (10pts).

Copy the Entity-Relationship diagram from Question 2 of HW4. Edit/update if necessary.



# Question 3 (40pts).

Given your current database model, list the following information about the imported data:

* Location of the available data;
* Size (in Mb, Gb, or the number of records) of the data.
* Mode for acquiring it (ftp download, screen scrape, request from authors or publications, publication holding servers, etc.)

NCBI journal results of animals from a zoo (emailed results can be filtered for data entry):

Location: <https://pubmed.ncbi.nlm.nih.gov/?term=%22Animals%2C+Zoo%22%5BMesh%5D&size=200>

Size: 1,444KB

Acquire: Scrape results from NCBI/Email results in a text file

Data scrape from HTML of online zoo website:

This website provides all the animals from a zoo from your location picking.

Location: <http://theonlinezoo.com/pages/north_carolina_zoo.html>

Size: 21 KB

Acquire: HTML Document that can be parsed with Python or R.

Animal classification and test life span information:

<https://genomics.senescence.info/species/index.html>: animal aging and life history website

Location: <https://genomics.senescence.info/species/dataset.zip>

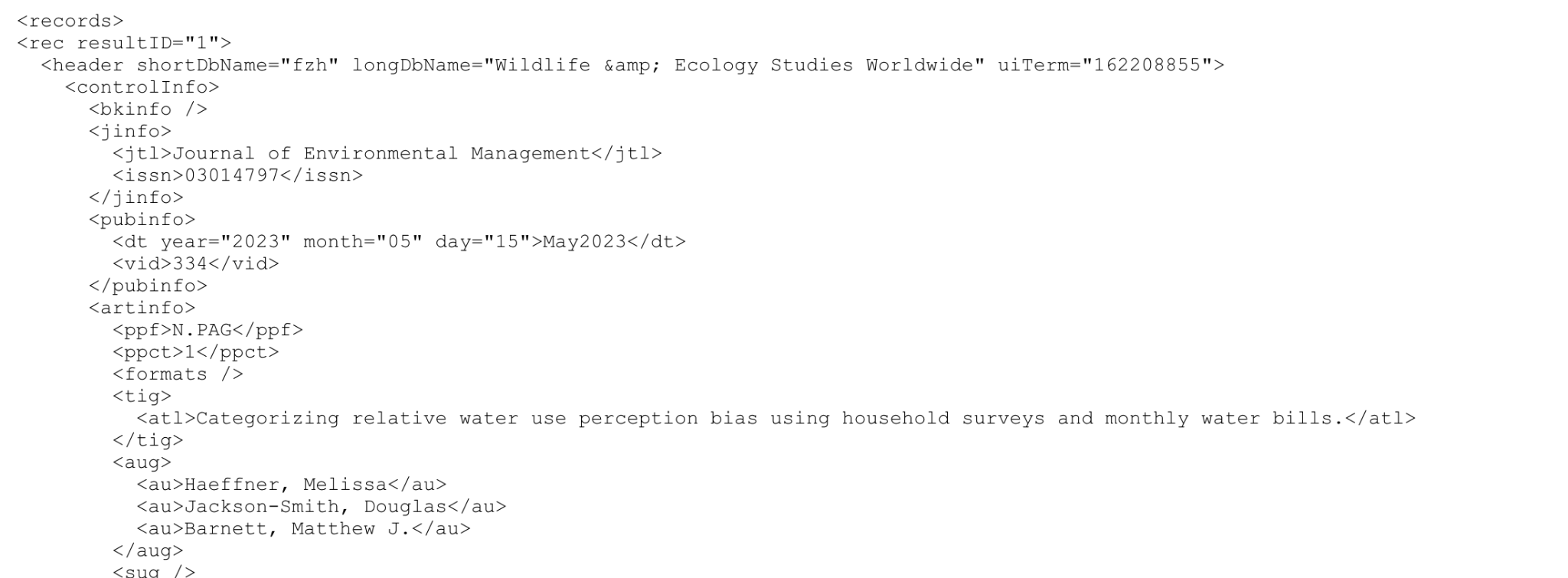
Size: 689 KB

Acquire: Txt file that can be read into Python, R, and SQLiteStudio directly.

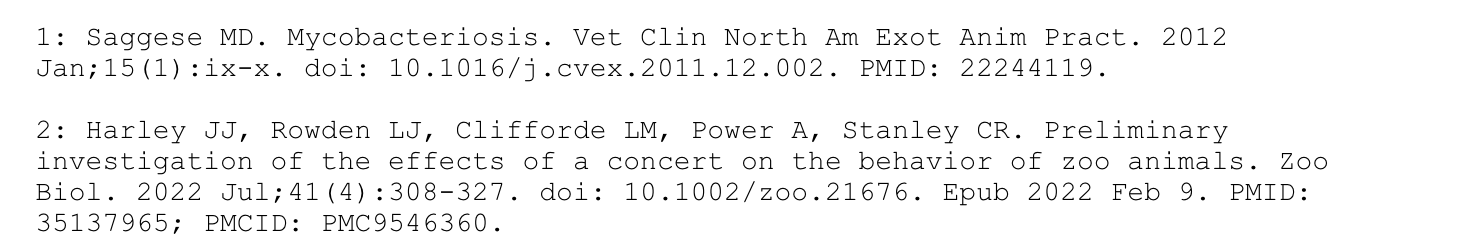
# Question 4 (40pts).

Provide a very small sample of the data files (you can attach them to your submission) and explain which fields will be imported from those files into your database.

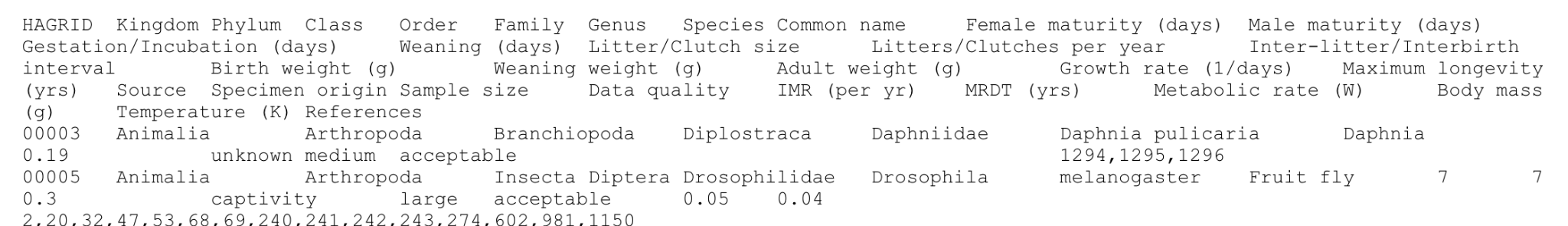
Animal Zoo Summary



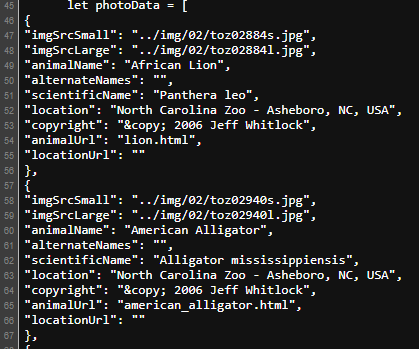
Animal Literature



Animal Classification



Zoo animal entry



For animal literature and animal zoo summary, all the fields provided can provide sufficient data for each literature provided. For animal classification, all fields of classification will be added along with HAGRID identification numbers. More information can be decidedly added in future database iterations, should they prove useful to connect future implementations. And finally for the zoo animal entry, the animal name, alternate names, scientific name, and location will be used to provide real online data provided from zoos.