## GIS 5572 Lab 1

**Due:** 2 weeks from the date of assignment

**Goals**

1. Practice decomposing interfaces for spatial web API’s into informal conceptual models
2. Compare and contract different web API’s using informal conceptual models and custom-built extract, transform, and load (ETL) routines
3. Build an ETL pipeline with Open Source Tools in Esri’s Online and ArcPro Jupyter Notebook (Can’t Find?)

**Deliverables**

Submit a lab report on Canvas as a PDF (see [report form](https://docs.google.com/document/u/0/d/1gOGBtTe3dQzrXCEMl644QIVdJgMp8ahN/?rtpof=true&usp=drive_fs)). Include all your code on Github.

**Specifics**

For this lab, write a lab report that does two things:

1. Compare and contrast the conceptual models for the following API’s
   1. [Minnesota Geospatial Commons](https://gisdata.mn.gov/content/?q=help/api) (CKAN)
   2. [Google Places](https://developers.google.com/places/web-service/overview)
   3. [NDAWN](https://ndawn.ndsu.nodak.edu//)
2. Create Jupyter notebooks that can programmatically get data from each of these APIs. Make all of this code available on Github in your Lab 1 folder.

A few tips:

1. Before writing any code, start by using paper and pencil to unpack the dataset objects.
2. Look at other examples of how people designed ETL code.
   1. Towards Data Science [article](https://towardsdatascience.com/integrate-jupyter-into-your-data-pipeline-9a02fab3cee5) on ETL with CRON or Jupyter
      1. Google terms you don’t understand (there are a lot of resources)

**Lab Report**

Title: Lab 1

Notice: Dr. Bryan Runck

Author: Cole Anderson

Date: 2/10/21

**Project Repository:** *<weblink to public repository>*

**Abstract**

*<Delete this text in light grey throughout>*

*250 words max. Clearly summarize the following major sections. Each gets one or two sentences.*

**Problem Statement**

*Describe the specific problem and the context. Provide an illustrative figure and/or context map here. In the table, translate the qualitative problem statement elements into specific requirements for the analysis.*

*Table 1. <insert caption>*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **Spatial Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |

**Input Data**

*Describe the data in two paragraphs max. Fill out the table.*

*Table 2. <insert caption>*

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
|  |  |  |  |

**Methods**

*Include a data flow diagram or screenshot from model builder. Do references in line (Rammankutty, 2033). Document any and all steps that you did to the input data in the data flow diagram. Provide natural language description of the most important steps, giving a narrative arc and provide well formatting screenshots with a boarder and centered throughout.*

**Conceptual Model, NDAWN**

**Method Flow, NDAWN**

**Conceptual Model, Places API**

**Method Flow, Places API**

**Conceptual Model, CKAN**

**Method Flow, CKAN**

*Resources on Data Flow Diagrams:*

* [*https://www.visual-paradigm.com/tutorials/data-flow-diagram-dfd.jsp*](https://www.visual-paradigm.com/tutorials/data-flow-diagram-dfd.jsp)
* [*https://www.lucidchart.com/pages/data-flow-diagram/how-to-make-a-dfd*](https://www.lucidchart.com/pages/data-flow-diagram/how-to-make-a-dfd)

*Figure 1. Data flow diagram.*

**Results**

*Show the results in figures and maps. Describe how they address the problem statement.*

*Follow best practice for map design, coloring, etc.*

What comes out of the API codes

**Results Verification**

*How do you know your results are correct? This can be a qualitative or quantitative verification.*

Checking API output vs. manual download

**Discussion and Conclusion**

*What did you learn? How does it relate to the main problem?*

**References**

*Use a common format*

Gette, C. (2018, May 18). Two (and a half) ways to get weather Data: Part 2- using an API and NOAA. Retrieved February 07, 2021, from <https://gettecr.github.io/noaa-api.html>

Google Maps Platform. (2020). Place search | places api | google developers. Retrieved February 07, 2021, from https://developers.google.com/places/web-service/search?hl=en\_US

Morales, F. (2020, May 20). Google maps: Places api. Retrieved February 07, 2021, from https://medium.com/swlh/google-maps-places-api-28b8fdf28082

Open Knowledge Foundation. (2012). The ckan api¶. Retrieved February 07, 2021, from https://docs.ckan.org/en/ckan-2.1.5/api.html

Prewitt, N., & Larson, S. M. (2020). Requests: Http for humans™¶. Retrieved February 07, 2021, from https://requests.readthedocs.io/en/master/

Ronquillo, A. (2020, November 07). Python's requests Library (guide). Retrieved February 07, 2021, from https://realpython.com/python-requests/#the-get-request

**Self-score**

*Fill out this rubric for yourself and include it in your lab report. The same rubric will be used to generate a grade in proportion to the points assigned in the syllabus to the assignment.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 |  |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 |  |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 |  |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 |  |
|  |  | 100 |  |