**Case Study – Summer 2018: Summary of Project Proposed by Dr. Scott – DRAFT**

**Team Members:** Michael Brush, Nicole, Vanessa Damato, Nicole Eversman, Lori Wickert

***“Working Title” –“The Use of WebScrapers and API’s as Tools for Populating a GeoSpatial Database for Missouri” (this may be too narrow)***

**Location of Study Area suggested:** State of Missouri.

* Many ‘known’ data sources available
* Additional sites on the web... whatever we come up with
* Access to additional data sources (non-public) may also be made available by Mizzou-DSA

**Objectives:**

What will we do?

* Connect, Assess, Visualize, Populate, Analyze and Interpret Geospatial Data for a defined study area from public and non-public data sources over the web.
* Use a web scraper to collect geospatial data from many widely available, ‘free’ data sources to populate a geospatial database over a study area.
* Web scraper for Wikimapia is already available (SQL snippet: cim\_schema.sql) – shared by Dr. Scott to all of us on slack.
* Connect other data sources as well – demonstrating how this can be done, via API’s or other solutions, compile and analyze the data to provide answers to our questions

**Particular challenges (opportunities) for our team:**

* Strong geospatial data background knowledge.
* Established coding skills.
* Remote sensing (satellite data) knowledge – data type, use and processing capabilities.
* Strong multidisciplinary background & analytical skills.

**Suggestions from Dr Scott:**

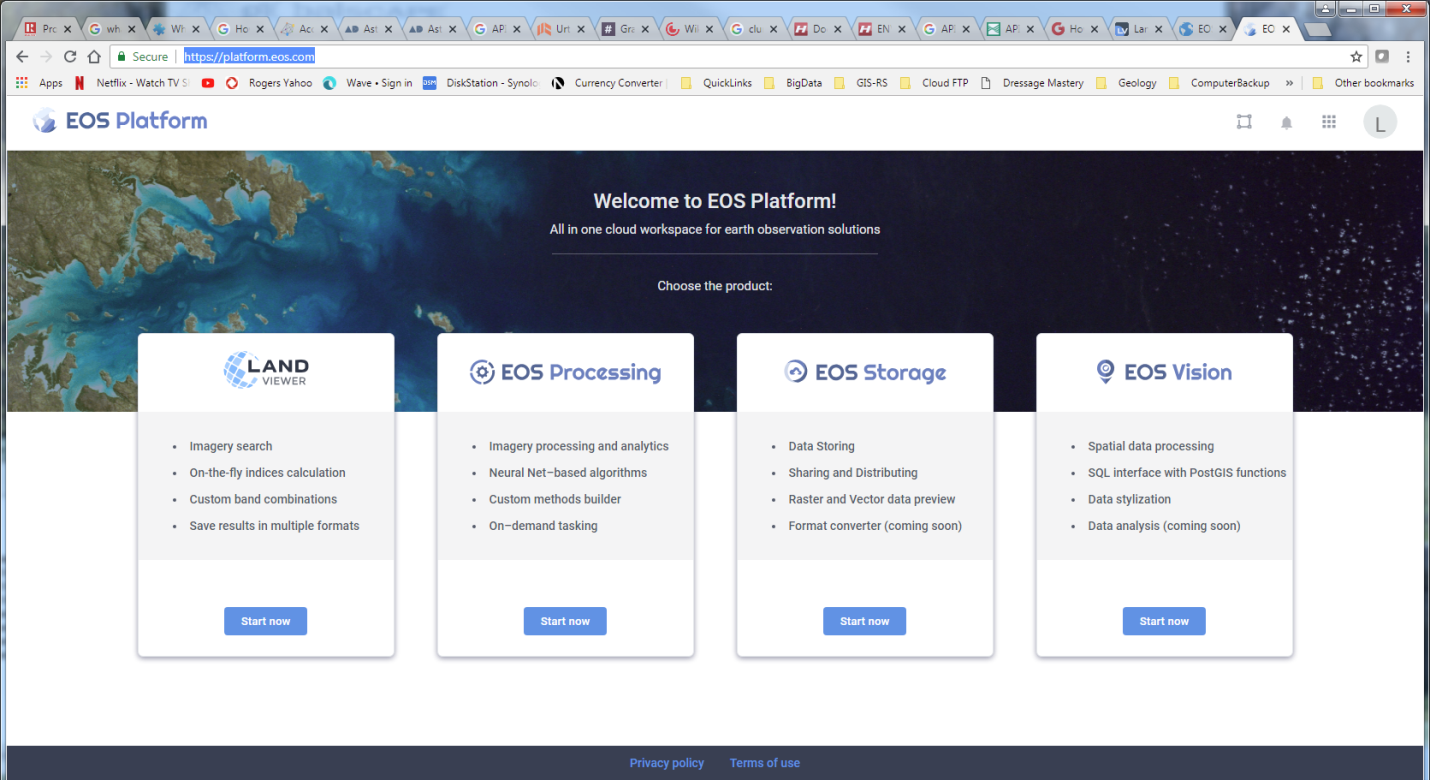
With the capabilities of our team, we are encouraged to... “Think outside of the box”... don’t worry about trying to define limits in scoping out what we can do – we don’t need to worry about overpromising. He wants us to think of this as a “no limits” project where we won’t be held accountable to what we say we would like to be able to do, but rather heralded for being innovative, making connections and links to data sources and accessing and linking things that maybe aren’t obvious or not something you would have planned. We don’t have to define how far we need to go with this project, no promises required, but rather encouraged to explore, adapt, be innovative and see how far we can take it...

**Question – data storage (?)**

* Needs to be cloud based... GIS(?)... Geospatial data functionality for sure ??

Git Hub (as suggested by Michael. Will this also work for geospatial data/what about visualization?

EOS Platform – would something like this also work for us? <https://platform.eos.com/>



**Next Steps:**

* Look at Wikimapia for Missouri. What kinds of data are available on that site? What questions can be asked with this kind of data... start to think if this is something we want to take on.
* Do some reading on API’s. Explore the API links for imagery.
* Assess this data based on the rubric for the case study. Would this data provide the information upon which specific questions could be answered? Is there sufficient breadth/depth in the data to do that?
* What role could you play in this process, your teammates? How can you best benefit from their skill sets and interests, as well as address their learning goals.
* Start to think about and even test access to the data/examine the data itself. Conduct a preliminary assessment of the data (ref: to Case Study description and rubric).

**Communication ?**

> We should establish a location (e.g. Google docs) where we could keep and update documentation for the team. Such as “hit list” like the above for data sites we have found, ideas and questions we have come up with, and connections we have working.

**Data Sources Identified To-date & links:**

*From Dr. Scott:*

**Wikimapea** - <http://wikimapia.org>

*Lori:*

**Geoawesomeness.com** – <http://geoawesomeness.com/>

* ‘Start thinking geospatial’ – what data is out there and what we could do with it... this is a great place to start to think outside of the box. This is also a place where we can start to formulate some questions (and potential answers) that could be derived from our project.

**Satellite Imagery & Other Earth Resource Data:**

***USGS (free) data sources***: <https://eros.usgs.gov/find-data>

***USGS Web Applications*** (kind of like our project): <https://www.usgs.gov/products/data-and-tools/web-application>

***Land Viewer:***<https://eos.com/landviewer/?s=Landsat8&id=LC08_L1TP_027033_20180509_20180517_01_T1&b=Red,Green,Blue,Panchrom&lat=38.34596&lng=-89.99451&z=7> Here is an example of easy, public access to ‘free’ imagery, which also has some ‘basic’ processing capabilities (see band combinations on right edge)

*Question* – how will imagery (e.g. geospatial data) provide solutions? Worth reading/place to start... http://geoawesomeness.com/how-digitalglobes-satellite-imagery-is-going-to-help-improve-openstreetmap/

**API – Data and Tools – USGS**: <https://www.usgs.gov/products/data-and-tools/apis>

There is a wealth of information here on available free data, data access, data management and tools. There is a whole list of existing API’s here. This could be a great source of API’s we can use to start to make connections to our data and get idea of what kind of data we want and how to use it. These are kind of like examples of the kind of project we are wanting to put together (I think).

**API’s/Access to Imagery:**

**UrtheCast** – have a demo API for ordering imagery – website with explanation of API’s and UrtheCast, and explaining the process by which the API works – very interesting.

<https://blog.urthecast.com/updates/earth-apis-ordering-geotiff-imagery/>

Demo: <https://github.com/urthecast/ordering-cli-demo>

**EOS Platform** – (as highlighted above): <https://platform.eos.com/>

**DigitalGlobe GBDx:** <https://platform.digitalglobe.com/gbdx/>

* This site provides access to developers to all of DG’s imagy archive (depending on your type of access). Free accounts are available, we may also be able to get special access if we are supported for our project via DG. Note that DG are also corporate sponsors to Mizzou-DSA.
* Look at some of the whitepapers here... see what is possible. Look at the kinds of solutions that people have come up with using this data source (e.g. James (Jimmy) Crawfod – Orbital Insights – I know him, and have had discussions about his project at the DG Consumer Advisory Board).
* DG also has an analytics capability onboard in their GBDx, including ENVI Tools, which is also supported via Harris. A description of this capability is provided here: <https://platform.digitalglobe.com/gbdx/imagery-analytics/>