01-06-2020 - BDH Meeting Notes

Attendees:

- Marcus Slavenas
- Jong Lee
- Shannon Bradley

Meeting Agenda:

- Review Project
- Outline Initial Project Tasks

Discussion:

What is Big Data Hub? Interoperability of systems and sharing of data between Iowa system and GLTG

Director Larry - IIHR

GLTG collect data and share to UMIS system

Consider it an extension of GLTG

Two aspects:

- Data sharing thru API - Geostreaming API v3 - not in clowder
  - So far they just want to use API
  - They will harvest data from us
- We propose we want to have python - jupyter notebook example of how to use API

Jong will connect Marcus to Team at University of Iowa

Marcus to do Immediately: list of data sources (send to Ibrahim and Jerry Mount)

Is there an end point to request? May need to create

Format: Excel or csv

/sensors

Choose which columns to export

Send all data points for all sensors

Decide on Column Titles:

ID	DS	Start Date	End Date	Parameter(s)

Send Jong list of column tomorrow

Next Marcus Task: We need proper documentation for API - URGENT - Deadline February 28

Use Open API spec - swagger

Open API Editor

Turns into JSON or yml file

Start with simple one first

- Geostreaming API documentation in OpenAPI Spec

Include Example returns / create a specification

Gowtham has used - openapi swagger

In-core Example - **Link? Jong Lee**

Share API with University of Iowa

If they start harvesting - and they harvest our datapoints from start to end they may choke our server

Put a limit on number of data points downloaded - API level - error only allow certain # points

Include in guideline and talk to them - do in chunks (human approach)

Current Documentation:

https://opensource.ncsa.illinois.edu/confluence/display/GEOD/Geostreaming+Api+V3+Deployment

Jong Lee - will make a new page to consolidate / re-organize

Phase 2: API Improvements after that based on their requests

Tied to Py-Geotemporal improvements

Risk Factors:

If they cannot use or do not approve our API:

Contingency is dump of database

Task - Jupyter Notebook on how to use API

Python

- Use py_geotemporal library to access
- Pandas or numpy
- Use in their analaysis framework