Geospatial Data Platform for All

Siddharth Ramavajjala, Sagi Zisman, **Ambarish Nag**, Travis Williams, Jianyu Gu, Duplyakin Dmitry

National Renewable Energy Laboratory, 15013 Denver West Pkwy, Golden, CO 80401.

NREL/PO-2C00-87748

Purpose

Spatial-temporal data has evolved in scale due to augmented use in cross-domain applications. **GIS** data presents unique challenges:

- Data discovery and access
- Efficient data storage
- Complex workflows requiring special expertise

Therefore, we propose a **laboratory-scale solution** to help domain and non-domain experts seamlessly store, analyze and incorporate spatial data into their analyses and applications.

Goals

Key aim is to contribute towards creation of a flexible and matured **geospatial data architecture**:

- Explore and design efficient GIS **data storage** solutions
- Improve **data discovery and access** supported through an interface
- Create a **Python library** to abstract common and complex **GIS workflows**
- Support geospatial big **data visualization** using Web Graphics Library (GL)
- Automate **data governance** of derived GIS products and support analyses

Data Platform Architecture

- Identify foundational datasets used across laboratory and map input data to processing requirements.
- Design independent databases to handle LiDAR and Raster data and build indexes to optimize data access and improve metadata-based querying.
- Data availability does not equate with accessibility. To improve data discoverability, a Knowledge-graph (**KG**) based discovery interface for semantic queries.
- Create a transformation library to package to abstract common GIS tools and operations, support data analysis.
- Co-locate storage & compute which triggers the complete pipeline execution using parallel compute. Provide data access through an easy-to-use API.
- Concrete data governance strategy to enrich the data platform through derived data and complement KG.

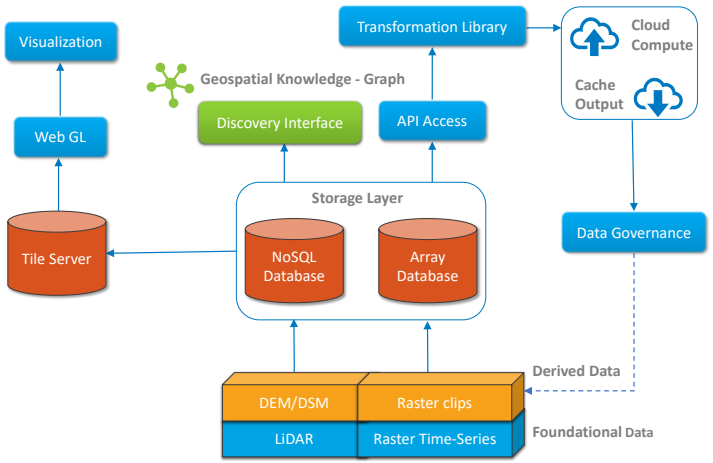

Figure 1: Sample architecture of data platform with raster and LiDAR foundational data

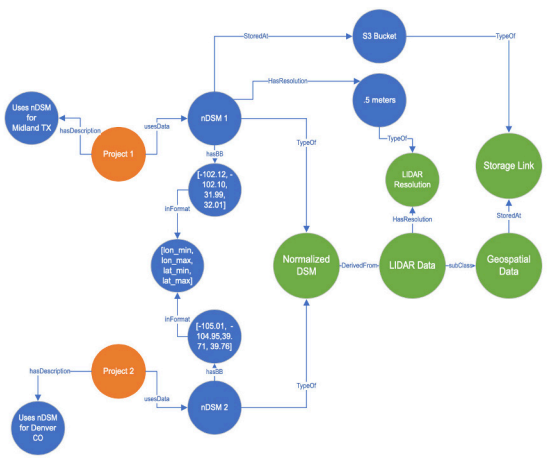

Figure 2: Sample implementation of KG over LiDAR data across two projects. Project 1 relates to Midland, TX location while Project 2 relates to Denver, CO location

- To overcome the data discoverability and the dataset linking issue, we propose a semantic KG-based data discovery service.
- KG allows, via graph traversal, for semantic data retrieval and querying without having to know a priori the dataset or its storage location.

Process

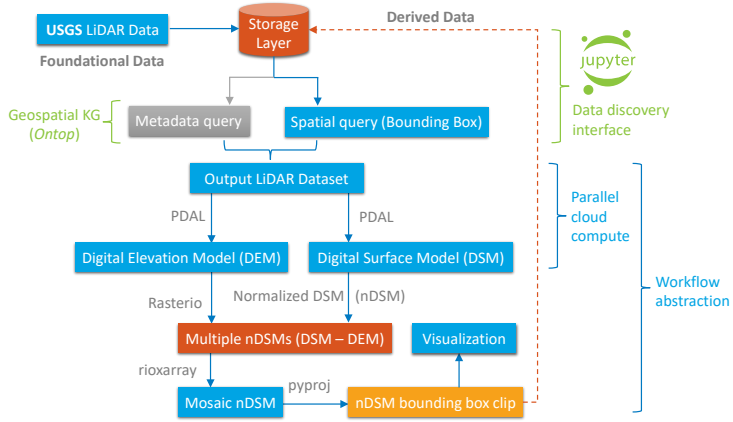

Figure 3: Sample processing steps to create nDSM from LiDAR data using simple bounding box coordinates query

- A single-line Python function abstracts complex workflow to create nDSM and masks GIS concepts, such as geo-query, interpolation, resolution and clipping, from the user.

```
# Diagonal coordinates to create bounding box
min_coords = [-102.07416294945199, 31.999611856879138]
max_coords = [-102.06987472552882, 32.001961043241579]

# Creates a nDSM and clips the extent to query area
get_ndsm_clip_from_geoquery(max_coords, min_coords, geoquery=1, interpolation=0, resolution=2.0)

Defined query bounding box has 1 LiDAR tiles available
filename: USGS_LPC_TX_West_Central_B4_2018_13RGR755430.laz is queried and retrieved to temp folder (164.6 MB)


Created DEM from LiDAR file  USGS_LPC_TX_West_Central_B4_2018_13RGR755430_dem.tiff
Created DSM from lidar_file  USGS_LPC_TX_West_Central_B4_2018_13RGR755430_dsm.tiff
Executed creation of nDSM USGS_LPC_TX_West_Central_B4_2018_13RGR755430_nDSM.tiff


Raster bbox_clip_102.07416294945199W31.999611856879138N_102.06987472552882W32.001961043241579N.tiff is clipped and c
reated to query bbox
```

User-friendly bounding box input

Figure 4: Python function that executes querying, processing of LiDAR datasets to create nDSM for the query bounding box

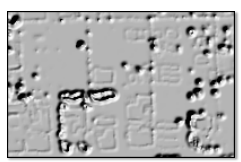

Figure 5: nDSM clip for query extent

Future Work

- Comprehensive preparation of foundational data, identification of critical application requirements and common GIS workflows through FY 24.
- Design of data discovery interface, implementation of an optimized geospatial KG, benchmarking databases for cloud deployment
- To develop and deploy an open-source transformation library and create a public-facing data platform to allow collaborators

Acknowledgement

- This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency. The views expressed in this poster do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.