Data Science for Plant Identification with Remote Sensing

Morteza Shahriari Nia and Dr. Daisy Zhe Wang

Data Science Research Lab

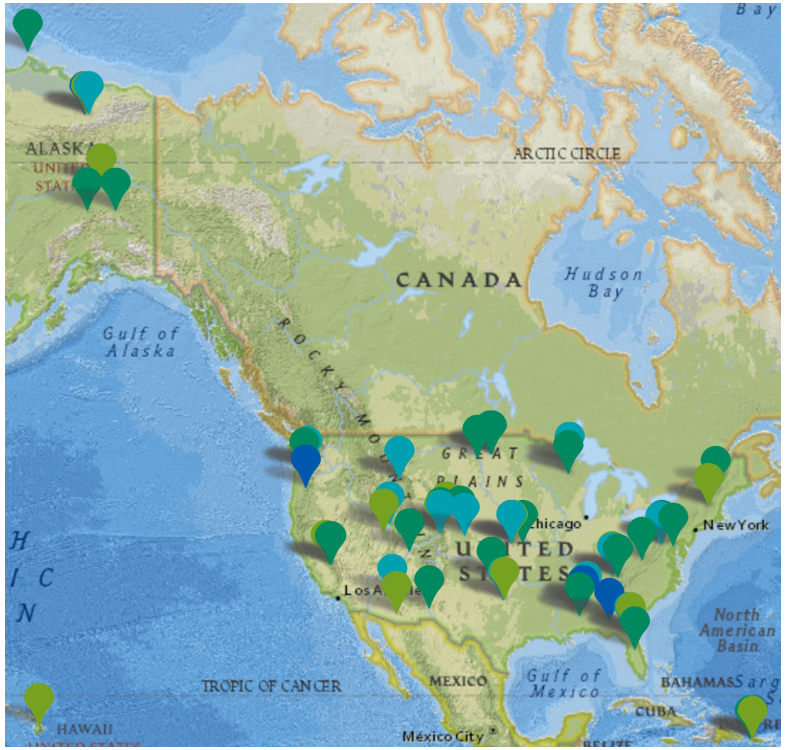
Department of Computer and Information Science and Engineering,

University of Florida

March 2016

Introduction

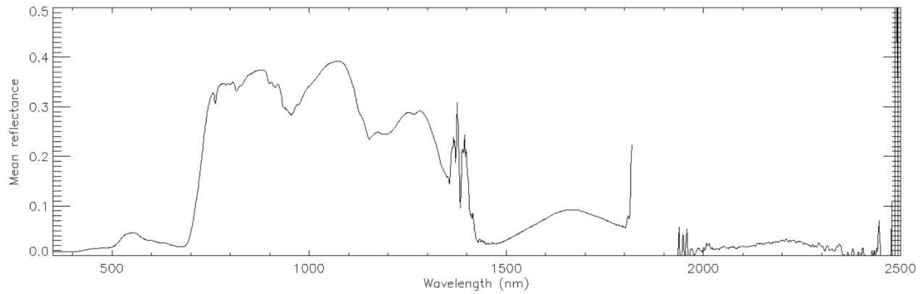
- Theme: Mapping tree species via airborne remote sensing at continental scale:
  - Land cover, climate change, invasive species, fire potentials, CO_2 emissions, soil characteristics
- Motivation: Field work vs remote sensing
- National Ecological Observatory Network
  - Funded by NSF, a 30-year project starting 2017
  - ~100 observatory sites
  - 100s of data products
    - Bioclimate
    - Biodiversity
    - Biogeochemistry
    - Ecohydrology
    - Infectious disease
    - Land use change
  - Routine and periodic data collection



Pilot Observatory Site

- Ordway-Swisher Biological Station
  - 37 km^2 - Putnam County in north-central Florida
  - Habitats: 9 habitats
- Species
  - 13 tree species + 1 class understory

Airborne Data Collection



- Airborne Observation Platform
  - Passive: Hyperspectral (12GB/flight, 10s flights per site)
    - Atmospheric Corrections
      - FLAASH (Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes)
      - ATCOR (ATmospheric CORrection)
    - 224 bands, wavelengths 400 nm to 2500 nm, resolution 3m

    ✧Challenges
      - Linear/nonlinear mixing of leaf, branch, soil, shade, and other signals
      - Timing of fights (September)
      - Type of trees (conifer vs broadleaf vs …)
  - Active: LiDAR (4GB/flight, 10s flights per site)
    - Point cloud

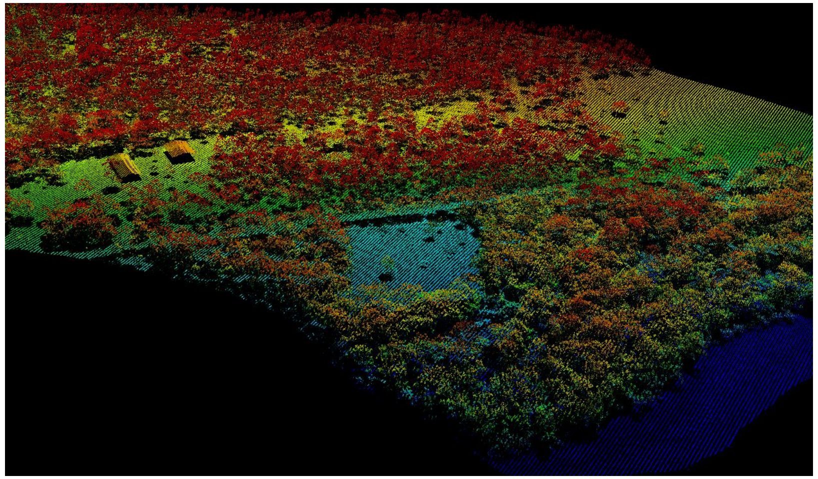


Figure 1. JPL AVRIS flights over OSBS [17]

(a) Flights ground tracks

(b) Hyperspectral true-color mosaic, morning 09/04/2010

(c) Hyperspectral true-color mosaic, mid-day 09/10/2010

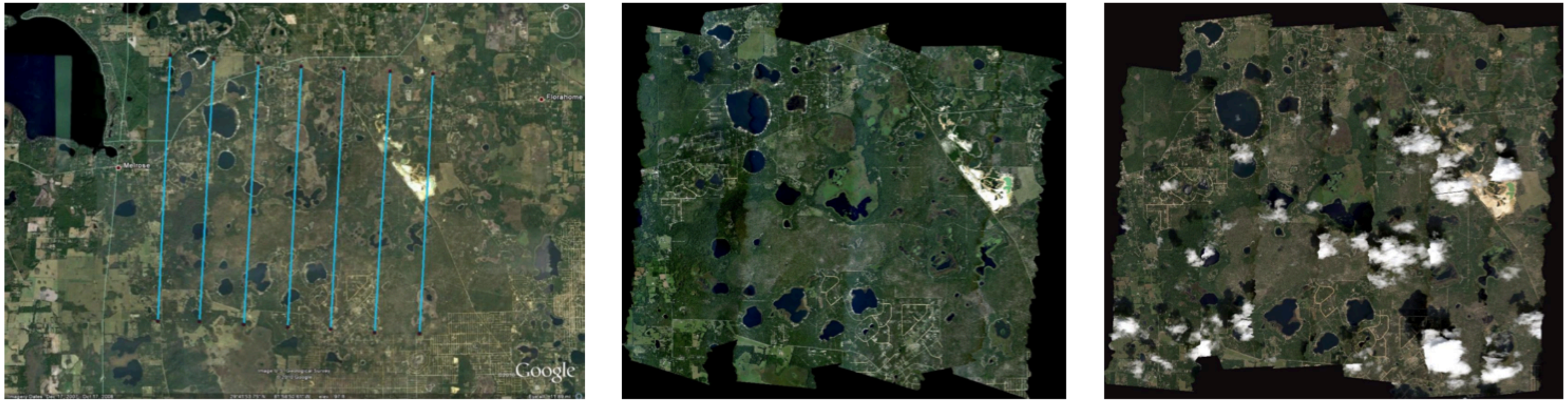
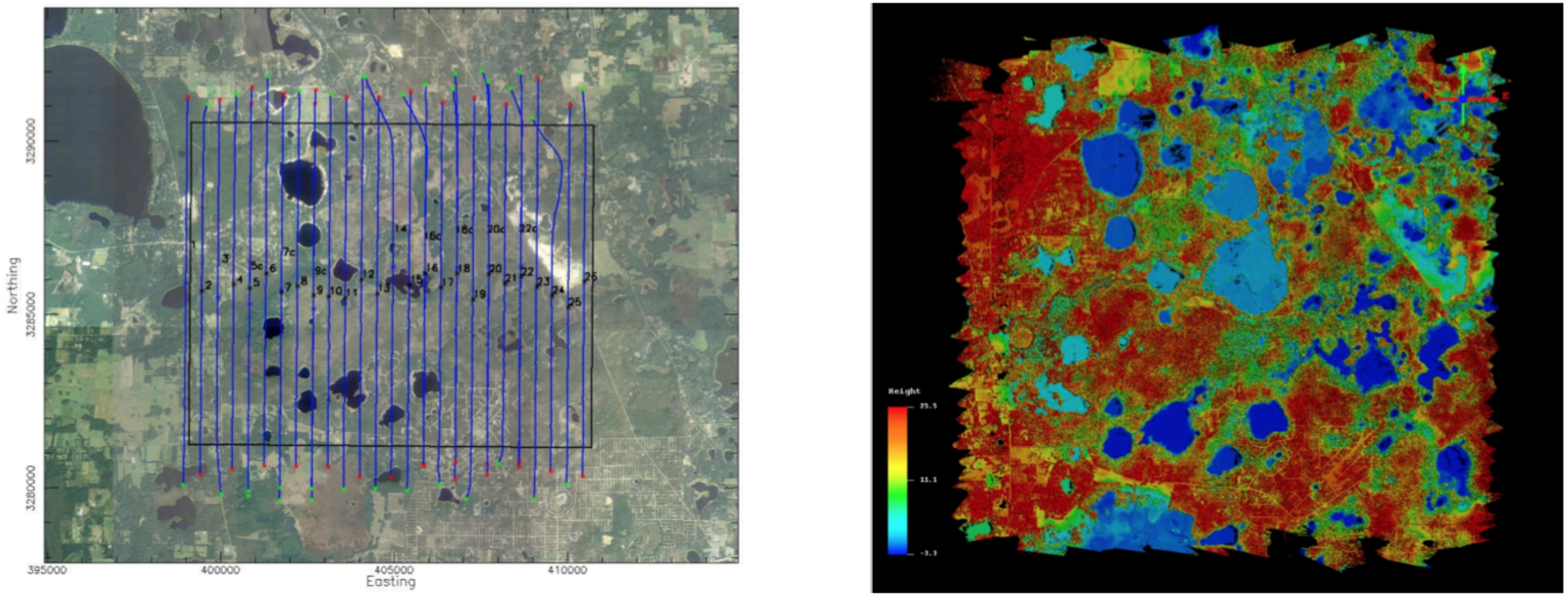


Figure 2. NCALM Optech Gemini LiDAR flights over OSBS 09/01/2010 [17]

(a) Flights ground tracks

(b) LiDAR color height mosaic



Field Data Collection

Data Types

- Raw (e.g. voltage)
- Calibrated (e.g. temperature, wind speed, radiance)
- Derived (e.g. NDVI, LAI)
- External (e.g. water evaporation, soil moisture, weather, land cover)
- Text (Structured/unstructured) (e.g. tables, charts, scientific papers, field guides)

A Data Science Challenge

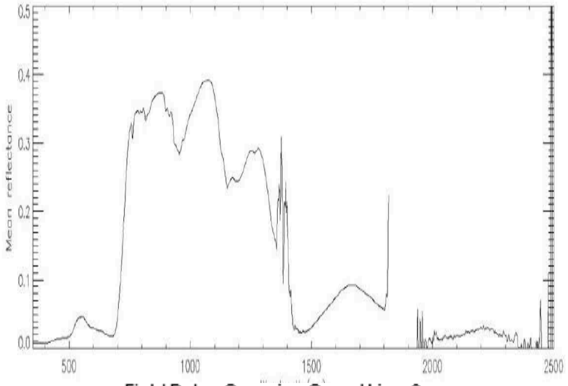
- Massive ecological model with precision details at continental scale
  - **Volume**
  - **Variety**
  - **Velocity**
  - **Value**

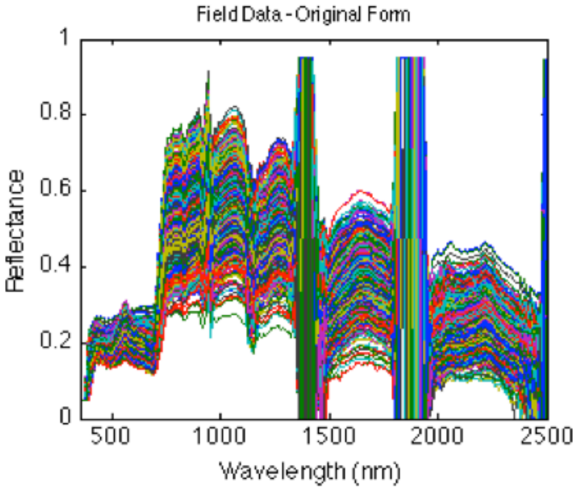
Task1: Data Cleaning

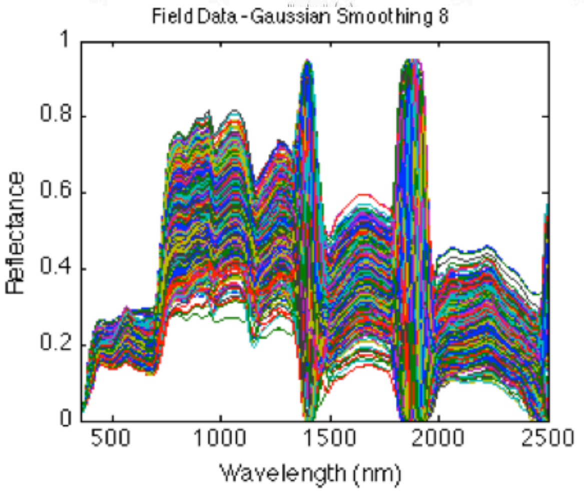
- Atmospheric correction
  - FLAASH
    - Physics-based algorithm from radiative transfer code
    - Eliminate atmospheric effects caused by molecular and particulate scattering and absorption from the radiance at the sensor and to obtain reflectance at the surface
  - ATCOR
    - Ground-based
      - Date, solar angle, visibility, elevation, etc
    - Flat/rugged terrain with Digital Elevation Model (DEM)

Task1: Data Cleaning (contd.)

- Remove noisy airborne data
- Adjust reflectance values
- Preserve low NDVI/NIR pixels
- Reduce cross-band sensor noise
- Reference signal: hand-held sensor
- Evaluation
  - Root mean squared error (RMSE)
    - Signal reconstruction from spectral library
  - Normalized mean square error (NMSE)







Task2: Aligning

- Equipment
  - ArcMap + ENVI + professional grade GPS
- Annotate polygons in ENVI image
- Limitations
  - Not possible in dense canopies
    - No-line of sight of GPS signals
  - GPS not accurate (off by a few meters)
    - Need landmarks
- Evaluation
  - Root mean squared error (RMSE)
  - Decision cost function (DCF)

Task 3: Classification

- Pixel species identification (stepping stone in ecology)
  - Understanding species unlocks doors for large amount of ecological research involving
    - plants volume
    - animals
    - insects
    - climate change
- Approaches
  - Use available variety of data
  - Machine learning models
  - Spectral library
- Evaluation
  - Precision, recall, F1 score, AUC

Spectral library: Multiple Endmember Spectral Mixture Analysis

- Linear unmixing based on abundance of each material in pixel

$$P'_i = \sum_{k=1}^{N} f_{ki} \times P_k + \epsilon_i, |f| = 1$$

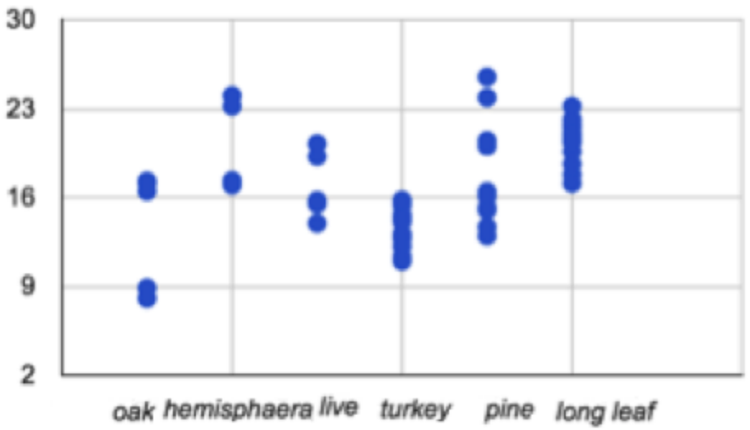
- Cons:
  - Very slow (combinatorial complexity)
  - Very inaccurate (about 50%) : highly dependent of library

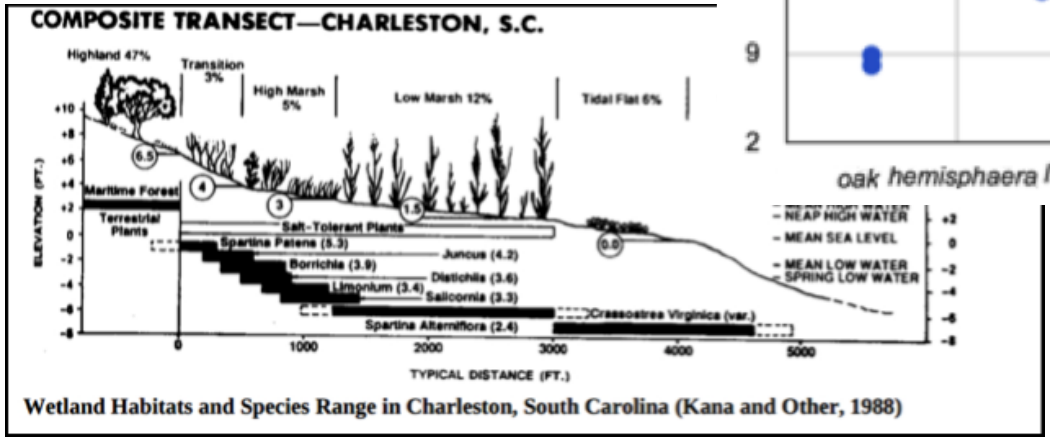
$$RMS = (||\epsilon_i||_2/N)^{1/2}$$

- Pros:
  - no need for massive field data!
  - Reusability cross-sites

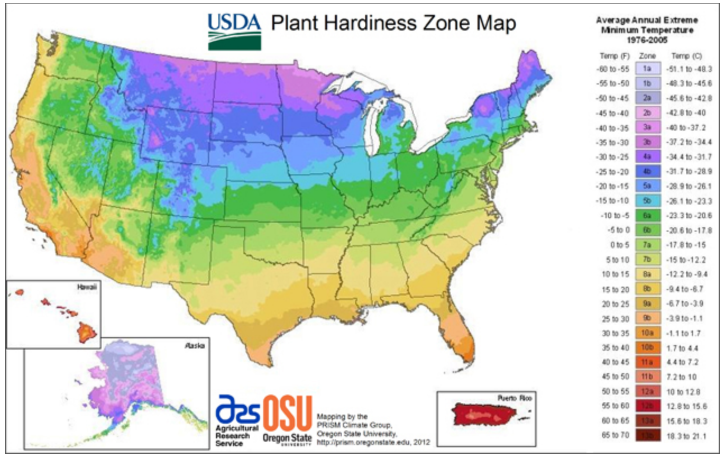
Knowledge Base for Ecology

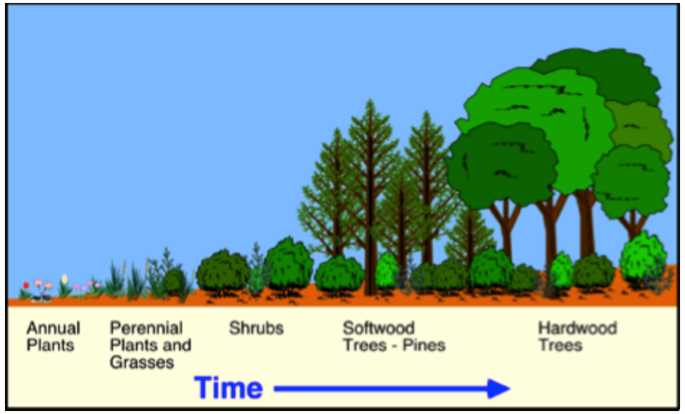
- 100,000+ species
- 1,100+ traits
  - Species cohabitation
  - Elevation
  - Height
  - Successional Level
  - Growth rate
  - Water stress tolerance
  - Cyclic changes
  - Plant Hardiness Zones
  - Precipitation
  - Shade tolerance





Wetland Habitats and Species Range in Charleston, South Carolina (Kana and Other, 1988)





Structured Data

- USDA PLANTS database

Conservation Plant Characteristics

***Quercus laevis* Walter**
turkey oak
QULA2

Morphology/Physiology

Active Growth Period	Spring and Su
After Harvest Regrowth Rate	
Bloat	None
C:N Ratio	High
Coppice Potential	Yes
Fall Conspicuous	Yes
Fire Resistant	No
Flower Color	Yellow
Flower Conspicuous	No
Foliage Color	Green
Foliage Porosity Summer	Moderate
Foliage Porosity Winter	Porous
Foliage Texture	Medium
Fruit/Seed Color	Brown
Fruit/Seed Conspicuous	Yes
Growth Form	Single Stem
Growth Rate	Rapid
Height at 20 Years, Maximum (feet)	30
Height, Mature (feet)	40.0
Known Allelopath	No
Leaf Retention	No
Lifespan	Short
Low Growing Grass	No
Nitrogen Fixation	None
Resprout Ability	Yes
Shape and Orientation	Erect
Toxicity	None

Growth Requirements

Adapted to Coarse Textured Soils	Yes

Other Data Sources

- EarthCube (funded by NEON)
  - Goal is to achieve interoperability and data integration across disciplines from computer science to Geo-sciences fields
- DataOne (funded by NSF)
  - Covers ecology, evolutionary, and earth science.
- iPlant/iAnimal

Other Datasets

- MODIS
- The Oak Ridge National Laboratory Distributed Active Archive Center
- NOAA Digital Coast Data Access Viewer (DEM)
- TRY datasets (land cover trends, water, wetlands, climate, …)
- Trait datasets TraitNet Plant traits – morphologi- cal, anatomical, physiological or phenological features mea- surable at the individual level
- landcover(forests/farms diversity and trends, aquatics and their effects)
- Alwyn H. Gentry Forest Transect Dataset
- Jornada Basin Long Term Ecological Research
- Major Data Hubs – with hundreds of submitted datasets:
  - http://ecologicaldata.org
  - http://datadryad.org/

A Challenge for Data Science

- Why NEON why now?
  - There is disparity between ecologist and remote sensing people. They don’t talk to each other. This proposal can be a big leap in marrying these two together as a data science challenge proposed by NIST.
  - Volume, variety, velocity → ecosystem dynamics (value)
    - Land cover, climate change, invasive species, fire potentials, CO_2 emissions, soil characteristics

Potential Participants

- Dr. Stephanie Bohlman (University of Florida)
- Dr. Ethan Whites (University of Florida)
- Dr. Paul Gader (University of Florida)
- Dr. Rob Guralnick (University of Florida)
- Dr. James Clark (Duke university)
- Dr. Timothy Keitt (University of Texas at Austin)
- Dr. Brian McGill (University of Maine)
- Dr. Brian Enquist (University of Arizona)
- Dr. Dar Roberts (University of California, Santa Barbara)

Questions?