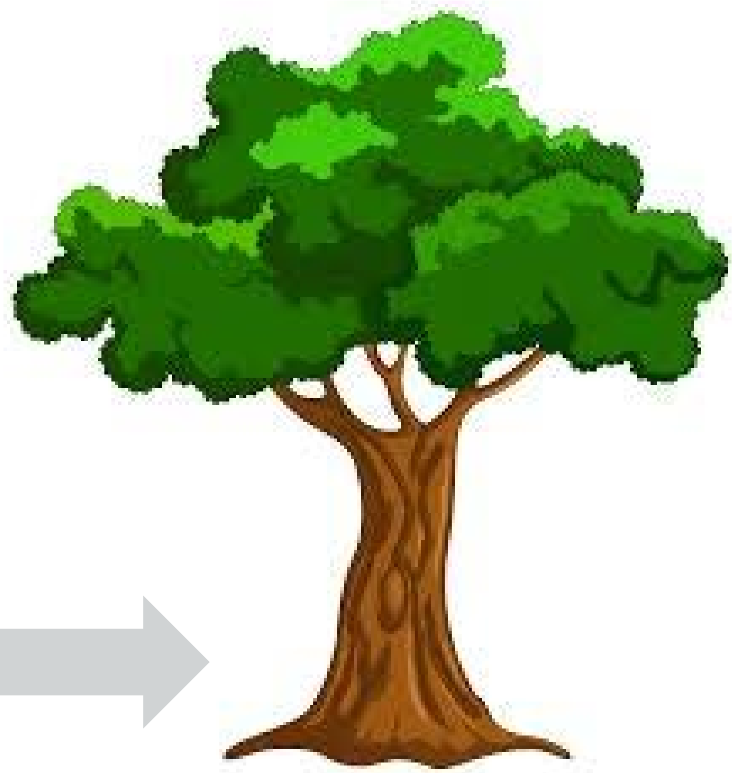
Jet Propulsion Laboratory
California Institute of Technology

Measuring and Predicting Data Value Using Scientific Literature

Kyle Hundman, Data Scientist

Measuring Value

- Nebulous, complex task
  - Publicity
  - Societal impact
  - **Science performed**
  - **Science enabled**



JPL

Focus on Literature

Science performed
Science enabled

Foundational to scientific discovery and research:

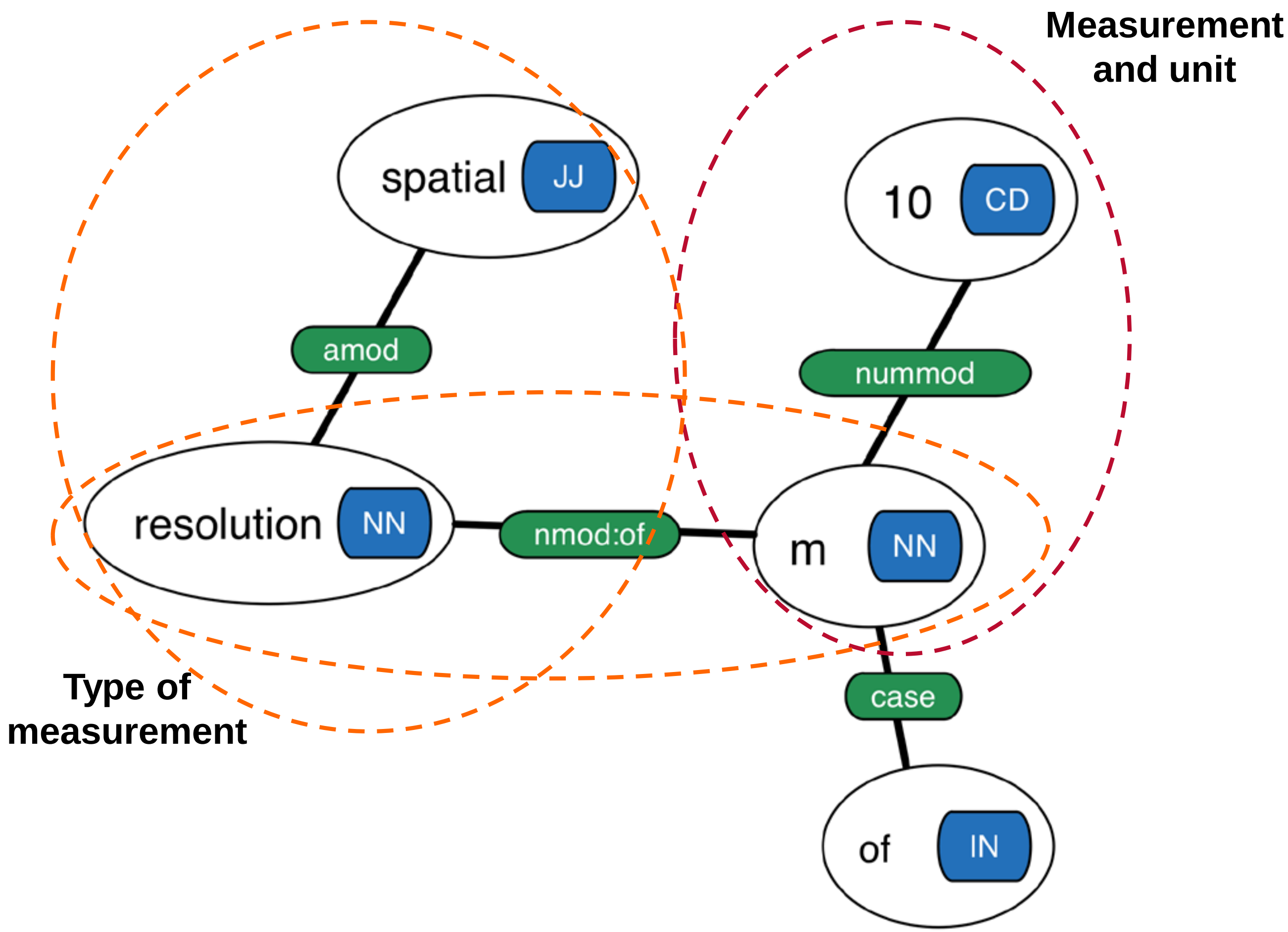
1. Literature
2. Proposals

One approach to linking is **data citations**

- Doesn't detail extent, nature of usage
- Impact of publication/research is unclear
- Descriptive, not predictive

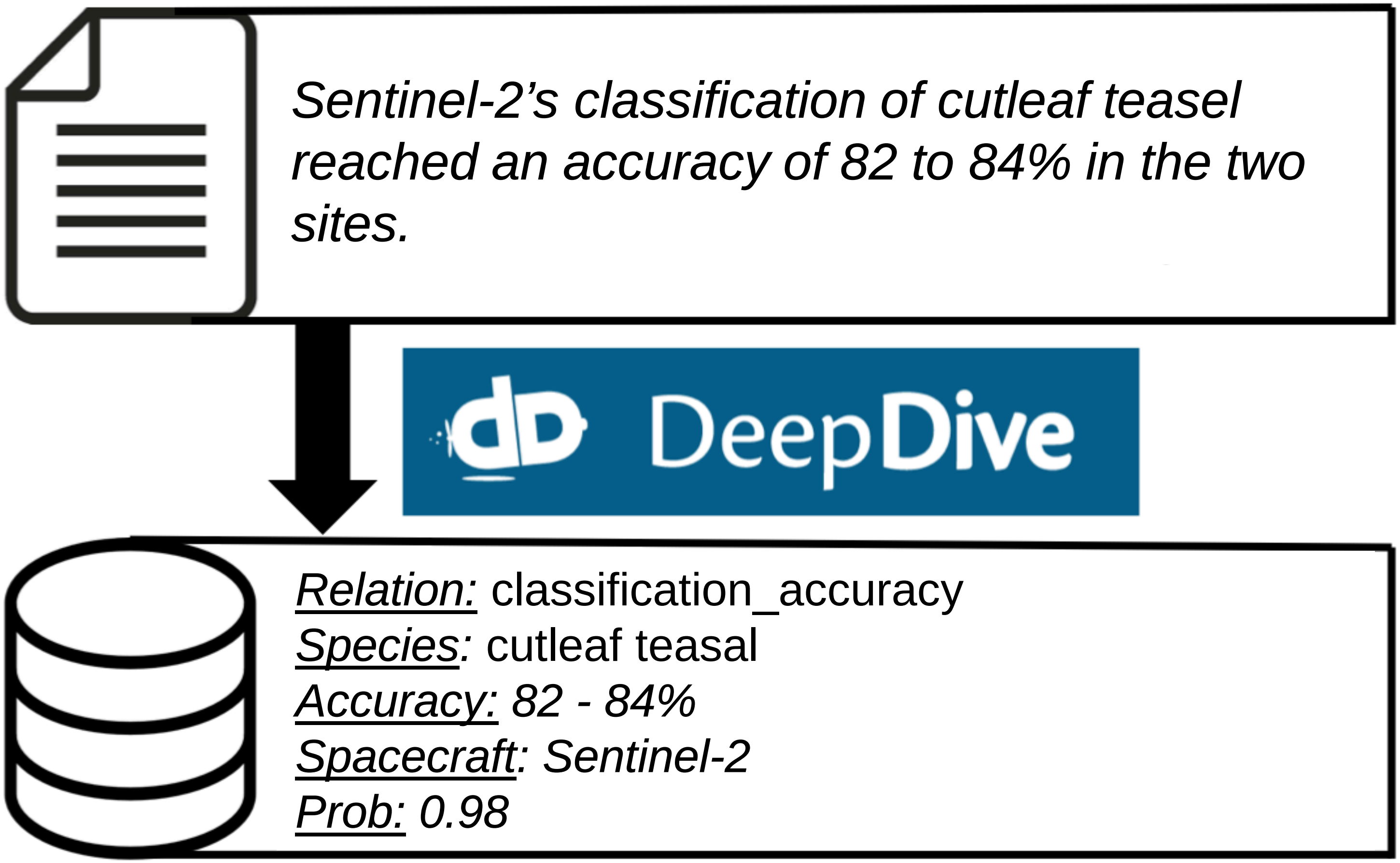
Measurement Extraction

Input: **"The spectrometer has a spatial resolution of 10 m."**



Output: { value: **10**, unit: **m**, type: **spatial resolution** }

Knowledge Base Construction



Start to answer questions:

- What were prior classification capabilities for cutleaf teasal?
- What is the impact to agricultural industry?
- Is there other data for this region that could improve this result?

HyspIRI Justification

- Mission justification in **Remote Sensing**
  - **HyspIRI** – hyperspectral imaging satellite

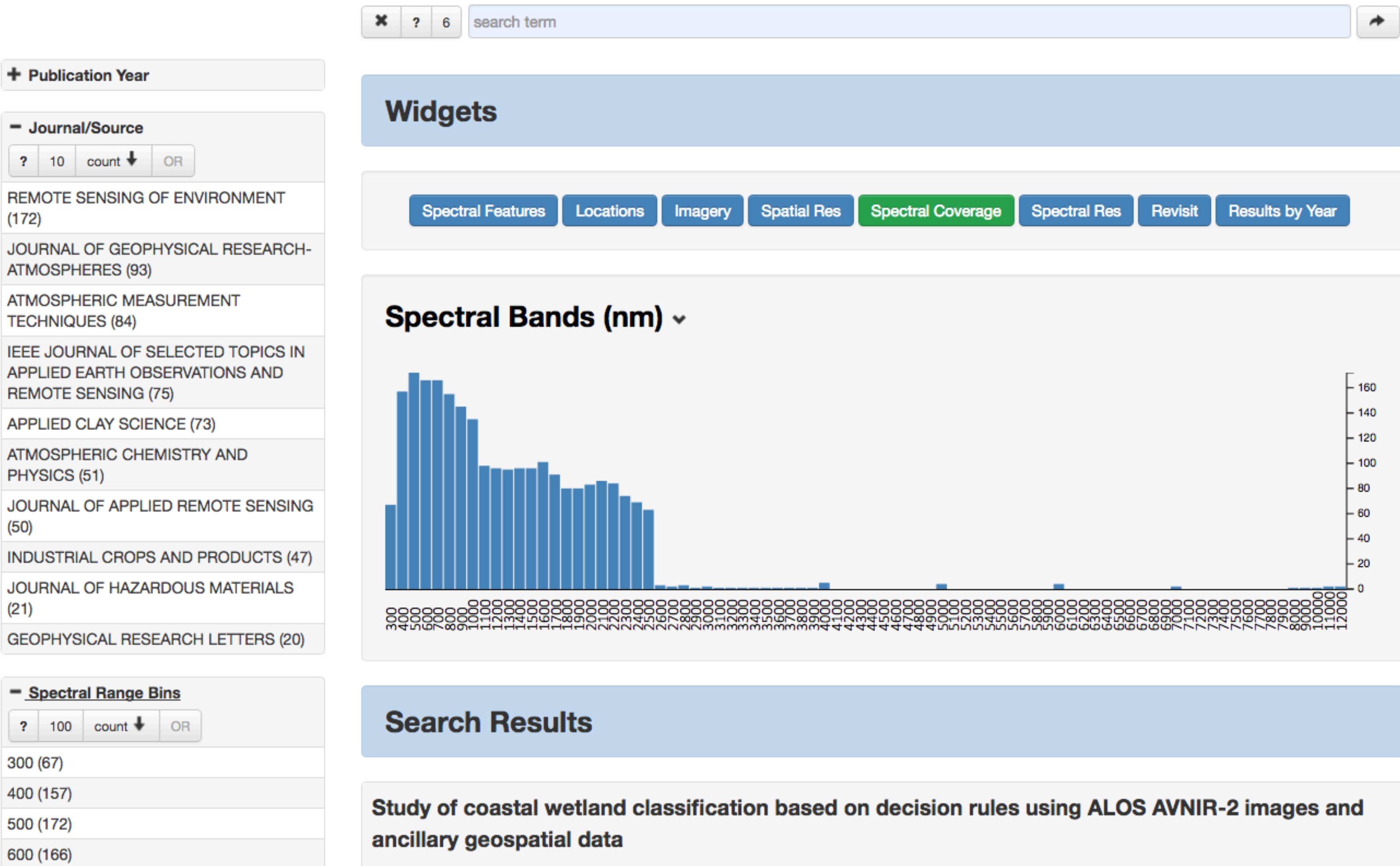
- Massive amounts of unstructured data
  - Web of Science search for “Remote Sensing” yields over **50,000 results**

- Extraction relevant information such as:
  - Measurements (e.g. spatial resolution)
  - Satellite mentions
  - Geophysical variables

- Search, visualize, analyze, structure heterogeneous data

Curated Search

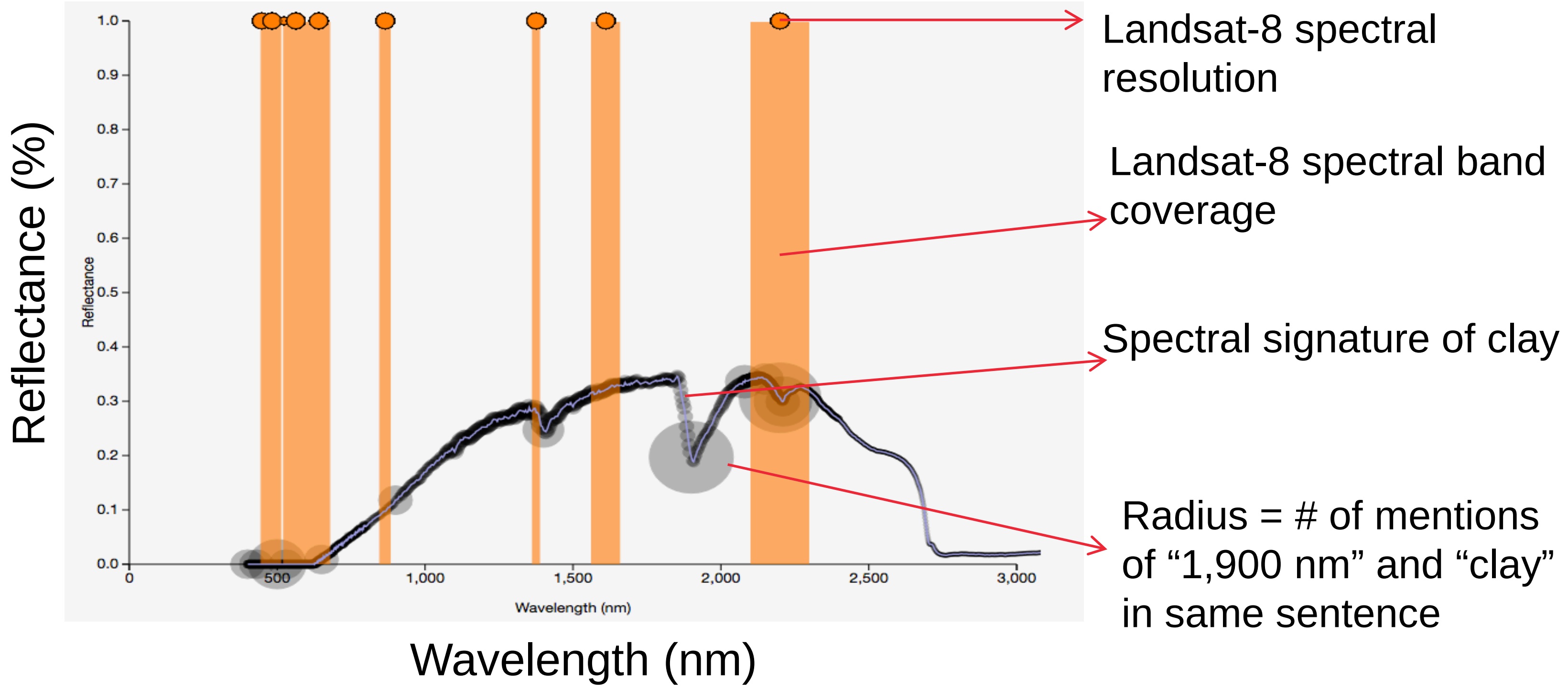
- Trends, outliers, exploratory analysis through faceted search



Assessing the Value of HyspIRI

- Looking for quantifiable justification for HyspIRI mission (top down)
    - **Measurement gaps**
    - Types of **scientific applications** enabled
- Parsed **2,500 journal publications** in remote sensing
    - joined with spectral reflectance data for various geophysical variables

- Example: { value: **1400**, unit: **nm**, related: [**spectral, clay, …**] }



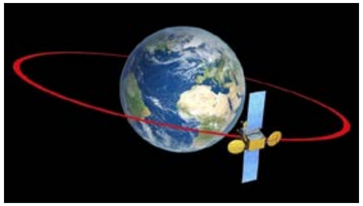
Decadal Survey for Earth Science

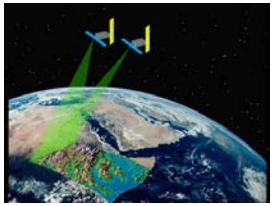
- “Generates consensus recommendations from the Earth and environmental science communities regarding a systems approach to space-based Earth Science observations.”
- Shapes science priorities and guides agency investments over the next decade
- Panel solicits input from community via two rounds of responses (RFIs)

Decadal Survey Motivations

- ~150 proposals
- Goal: Find **optimal mission architecture** using proposals

relationships between 3 facets of architecture:



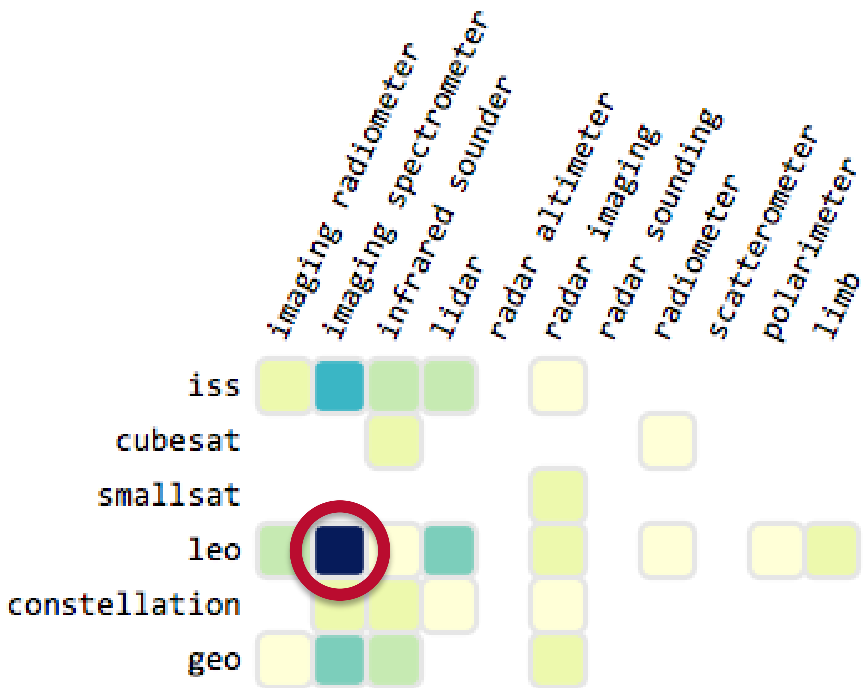


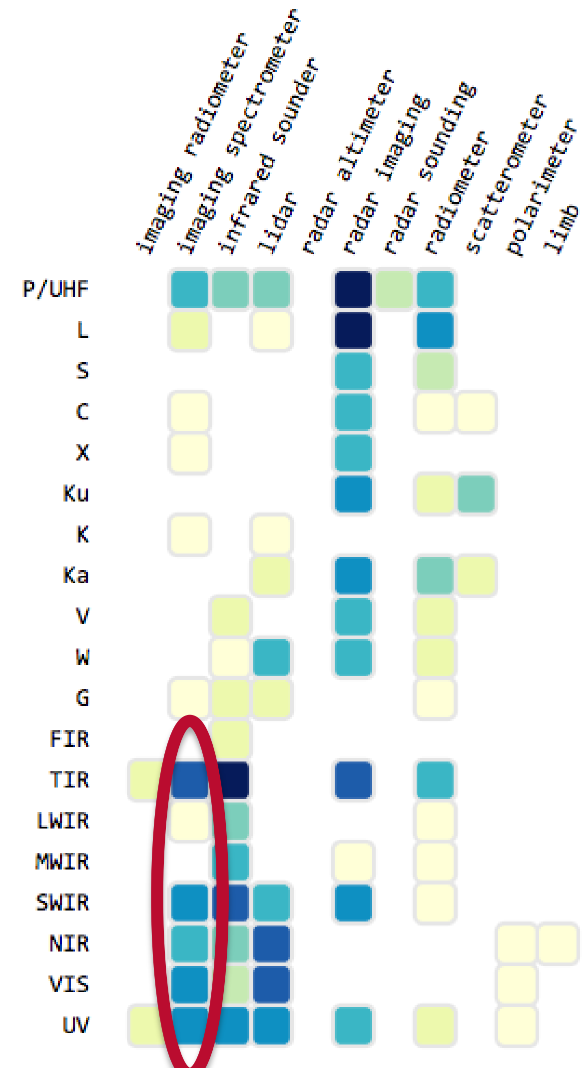
Orbit	**Instrumentation**	**Measurement Targets**
low earth orbit (LEO)	radar	soil moisture
geostationary (GEO)	lidar	aerosols
constellation (CubeSats)	IR Sounder	ice thickness

Search, Viz, Drill-Down

Structured data on relations between orbits, surveying methods, measurement targets:

RFIs, JPL subject matter experts focused on **infrared spectrometers** in **LEO**





Optimal architecture = maximum science

Architectures can achieve certain measurement targets
(infrared spectrometer in LEO is 1 of 2 architectures needed to measure **formaldehyde)**

Score = 1 / 2 = 0.5

Measurement targets can answer pressing science questions
(**formaldehyde** is 1 of 6 key components to understanding levels of **power production**)

Score = 1 / 6 = 0.17

LEO + Infrared spectrometer science value **= 0.5 * 0.17 =**

0.085

Results

1. More granular, comprehensive view of data usage

1. Expedite analysis and exploration of scientific results to better understand value

2. Move toward identification of opportunities and prediction of value

=

More Efficient and Informed Science

Links

https://github.com/khundman/marve
http://deepdive.stanford.edu/
https://github.com/chrismattmann/facetview

Jet Propulsion Laboratory
California Institute of Technology