Pipeline for Applications-Based Data Discovery

Katrina Virts (1), Prasanna Koirala (1), Iksha Gurung (1), Dai Hai Ton That (1), Muthukumaran Ramasubramanian (1), Brian Freitag (2), and Rahul Ramachandran (2)

(1) University of Alabama in Huntsville, Huntsville, AL; (2) NASA/Marshall Space Flight Center, Huntsville, AL

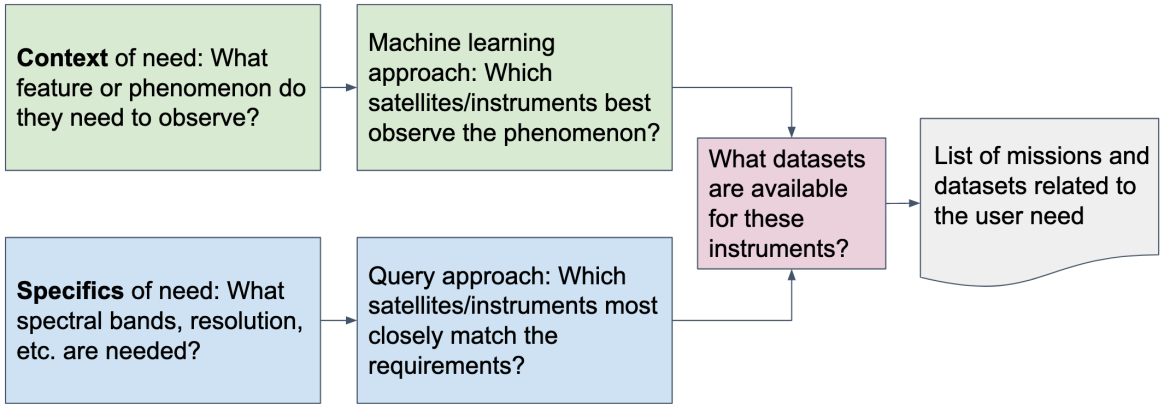
PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

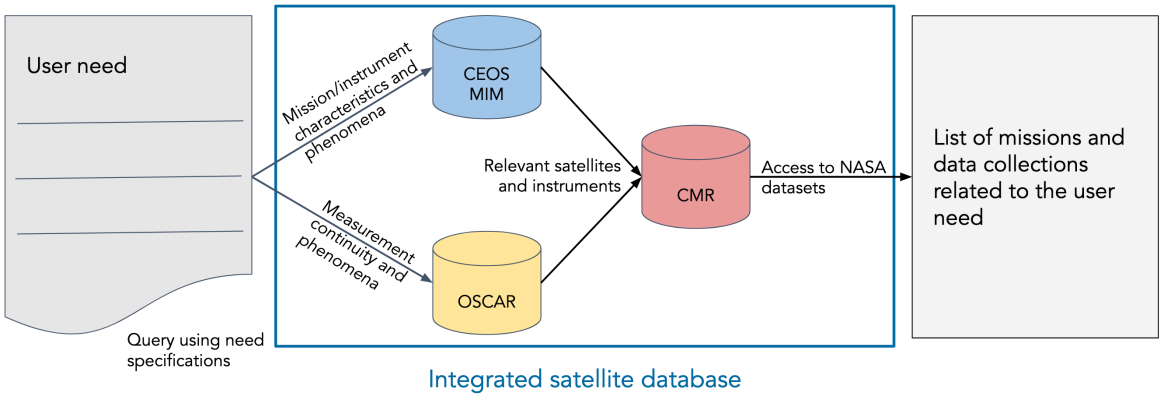
1. INTRODUCTION

From disaster response and mitigation to monitoring water quality or protecting wildlife habitat, satellite Earth observation data can be applied in countless ways to meet pressing needs and benefit society. **The crucial first step toward successful data application is data discovery.** Potential users often know exactly what data they need--what Earth feature or phenomenon they need to observe, how frequently, and at what resolution or level of accuracy--but may still struggle to discover the existing observations that meet their needs.

We have developed a pipeline to connect applications-based users to specific satellites and data collections within NASA's Earth observation program of record that are highly relevant to their data needs. This pipeline combines available information on satellite and instrument measurement characteristics with an innovative machine learning-based approach that identifies instruments that are most relevant to the feature or phenomenon of interest.



2. CREATING AN INTEGRATED DATABASE OF SATELLITES AND INSTRUMENTS



We have created a **Neptune graph database** that integrates information on Earth-observing satellites and instruments from two publicly available databases:

- **CEOS MIM** (http://database.eohandbook.com/database/missiontable.aspx) (Committee on Earth Observation Satellites Mission/Instruments/Missions) database
  - Contains past, current, and future Earth observation missions
  - Satellite specifics including: status, period of operation, orbit type, onboard instruments
  - Instrument specifics including: instrument type, phenomena/measurements, wavebands, spatial resolution, swath width, accuracy
- **WMO OSCAR** (https://www.wmo-sat.info/oscar/) (World Meteorological Organization Observing Systems Capability Analysis and Review Tool)
  - Additional information on phenomena and measurements of each instrument
  - Gap analyses showing the continuity of satellite observations of Earth features and phenomena

In addition, our pipeline accesses specific NASA data collections by using:

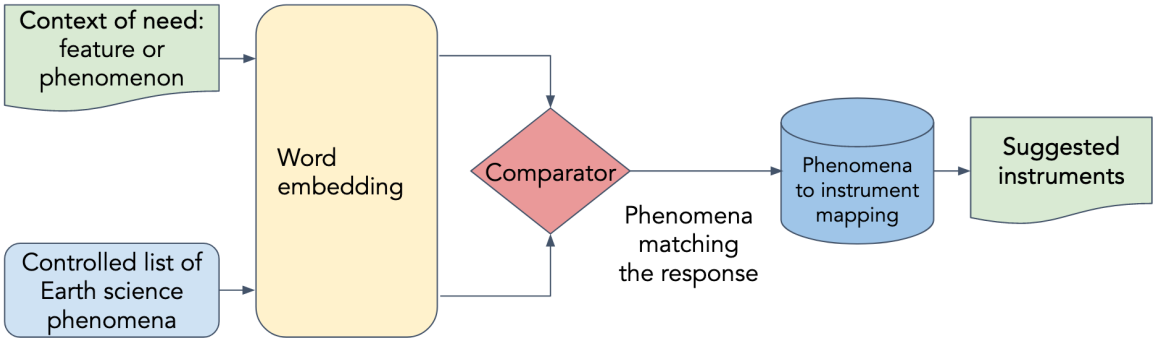
- **CMR** (https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr) (NASA's Common Metadata Repository)
  - High-quality repository that catalogs all data and service metadata records for NASA's Earth Observing System Data and Information System (EOSDIS)
  - Contains information about specific collections (datasets) produced from NASA instruments and some partner agencies
  - CMR enables Earthdata search (https://search.earthdata.nasa.gov/search) and can be queried using API

3. USING MACHINE LEARNING TO IDENTIFY RELEVANT INSTRUMENTS

What if the phenomenon the user needs to observe isn't in the list of instrument-related phenomena and applications in our database? **We use a machine learning approach based on *word embedding***, which:

- Converts words and phrases into a numerical representation based on their linguistic context
- Useful for performing statistical analysis on a corpus (large body of text)
- Word2Vec: two-layer neural network, captures linguistic and sematic relationships between words in a corpus
  - We have developed ES2Vec by training on a corpus of ~90,000 AGU journal articles

By comparing word embeddings for the user-provided phenomenon and the controlled list in the integrated satellite database, **we identify satellites and instruments that are closely related to the phenomenon**:



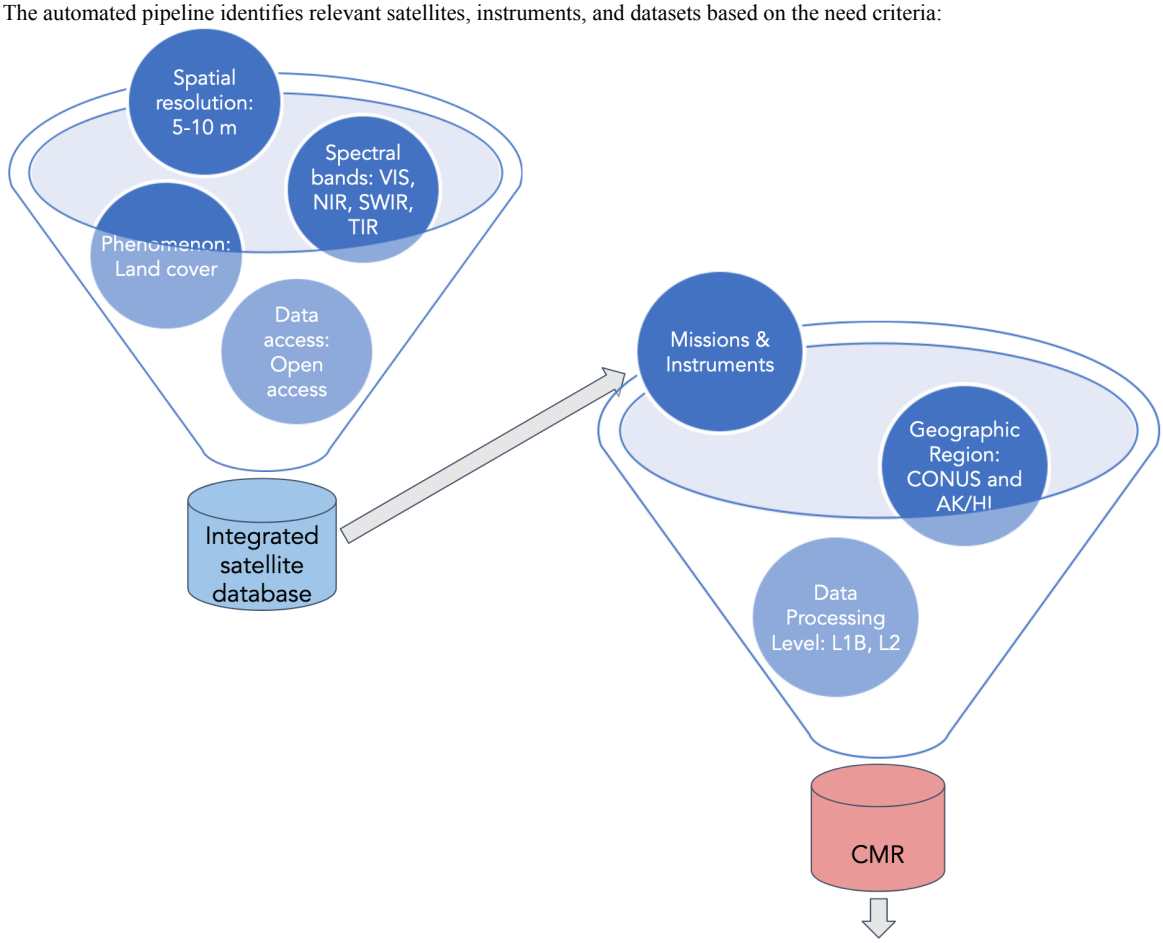
4. EXAMPLE OF THE PIPELINE IN ACTION

The applications-based data discovery tool will eventually be publicly available. It is currently being tested on a database of Earth observation needs submitted by federal agencies to the Satellite Needs Working Group (https://earthdata.nasa.gov/esds/impact/snwg).

Example agency need:

- Feature/phenomenon: Land cover
- Geographic domain: Continental United States and Alaska/Hawaii
- Spatial resolution: 5-10 m
- Frequency of observation: Multiple times per week
- Spectral bands: Visible (VIS), near infrared (NIR), shortwave infrared (SWIR), thermal infrared (TIR)

The automated pipeline identifies relevant satellites, instruments, and datasets based on the need criteria:



Candidate list:

Mission/ Instrument/ Agencies	Status	Criteria Matched	Collections
ICESat-2 ATLAS NASA	Operational (2018)	*Spatial resolution:* 6 footprints 13 m in diameter separated by ~1.5 cm along track	https://dx.doi.org/10.5067/ATLAS/ATL02.003 https://dx.doi.org/10.5067/ATLAS/ATL03.003 https://dx.doi.org/10.5067/ATLAS/ATL04.003
		Waveband categories: VIS *Feature/ phenomenon*	
Sentinel-2 A MSI (Sentinel-2) ESA, COM	Operational (2015)	*Spatial resolution:* 10 m for 4 bands in VNIR, 60 m for 3 dedicated atmospheric correction bands, 20 m for remaining bands *Waveband categories:* SWIR, VIS *Feature/ phenomenon*	https://dx.doi.org/10.5067/HLS/HLSS30.015
Sentinel-2 C MSI (Sentinel-2) ESA, COM	Approved (2021)	*Spatial resolution:* 10 m for 4 bands in VNIR, 60 m for 3 dedicated atmospheric correction bands, 20 m for remaining bands *Waveband categories:* SWIR, VIS *Feature/ phenomenon*	
Sentinel-2 B MSI (Sentinel-2) ESA, COM	Operational (2017)	*Spatial resolution:* 10 m for 4 bands in VNIR, 60 m for 3 dedicated atmospheric correction bands, 20 m for remaining bands *Waveband categories:* SWIR, VIS *Feature/ phenomenon*	https://dx.doi.org/10.5067/HLS/HLSS30.015

		phenomenon	
Aqua MODIS	Operational (2002)	*Horizontal resolution:* Cloud cover: 250 m (day)	https://dx.doi.org/10.5067/MODIS/MYD021KM.NRT.061 https://dx.doi.org/10.5067/MODIS/MYD021KM.061
NASA, JAXA, INPE		and 1000 m (night), Surface temperature: 1000 m *Waveband categories:* VIS, NIR, SWIR, MWIR, TIR	https://dx.doi.org/10.5067/MODIS/MYD14.006 https://dx.doi.org/10.5067/MODIS/MYD09.NRT.006 https://dx.doi.org/10.5067/MODIS/MYD09.006 https://dx.doi.org/10.5067/MODIS/MYD11_L2.NRT.006 https://dx.doi.org/10.5067/MODIS/MYD11_L2.006 https://dx.doi.org/10.5067/MODIS/MYD21.006
Terra MODIS NASA, METI, CSA	Operational (1999)	*Horizontal resolution:* Cloud cover: 250 m (day) and 1000 m (night), Surface temperature: 1000 m *Waveband categories:* VIS, NIR, SWIR, MWIR, TIR	https://dx.doi.org/10.5067/MODIS/MOD021KM.NRT.061 https://dx.doi.org/10.5067/MODIS/MOD021KM.061 https://dx.doi.org/10.5067/MODIS/MOD11_L2.NRT.006 https://dx.doi.org/10.5067/MODIS/MOD11_L2.006 https://dx.doi.org/10.5067/MODIS/MOD14.006 https://dx.doi.org/10.5067/MODIS/MOD21.006 https://dx.doi.org/10.5067/MODIS/MOD21A1D.006 https://dx.doi.org/10.5067/MODIS/MOD21A2.006
GOMX4 HyperScout ESA, NSO	Operational (2018)	*Waveband categories:* VIS *Feature/ phenomenon*	

VENµS VSC CNES, ISA	Operational (2017)	*Spatial resolution:* 5.3 m spatial resolution with 27 km swath *Waveband categories:* VIS, NIR *Feature/ phenomenon*	
RESOURCESAT-3 ALISS III ISRO	Approved (2021)	*Spatial resolution:* 20 m, 10 m *Waveband categories:* SWIR, VIS, NIR	
RESOURCESAT-3A ALISS III ISRO	Approved (2022)	*Spatial resolution:* 20 m, 10 m *Waveband categories:* SWIR, VIS, NIR	
RapidEye MSI DLR	Operational (2008)	*Spatial resolution:* 6.5 m *Waveband categories:* VIS, NIR	
ALOS-2 CIRC JAXA	Operational (2014)	*Waveband categories:* TIR *Feature/ phenomenon*	
Landsat 7 ETM+ USGS, NASA	Operational (1999)	*Waveband categories:* SWIR, VIS, TIR, NIR *Feature/ phenomenon*	
Pleiades 1B HiRI CNES	Operational (2012)	*Waveband categories:* VIS, NIR *Feature/ phenomenon*	
Pleiades 1A HiRI CNES	Operational (2011)	*Waveband categories:* VIS, NIR *Feature/ phenomenon*	
CARTOSAT-2E HRMX ISRO	Operational (2017)	*Waveband categories:* VIS, NIR *Feature/ phenomenon*	
RESOURCESAT-3S APAN ISRO	Approved (2020)	*Waveband categories:* VIS *Feature/ phenomenon*	
RESOURCESAT-3SA APAN ISRO	Approved (2021)	*Waveband categories:* VIS *Feature/ phenomenon*	
RESOURCESAT-2 AWiFS ISRO	Operational (2011)	*Waveband categories:* SWIR, VIS, NIR *Feature/ phenomenon*	
RESOURCESAT-2A AWiFS ISRO	Operational (2016)	*Waveband categories:* SWIR, VIS, NIR *Feature/ phenomenon*	
CARBONITE-2 CARBONITE-2 Imager UKSA	Operational (2018)	*Waveband categories:* VIS *Feature/ phenomenon*	

Suomi NPP VIIRS NASA, NOAA	Operational (2011)	*Horizontal resolution:* 400 m – 1.6 km *Waveband categories:* VIS, NIR, SWIR, MWIR, TIR	https://dx.doi.org/10.5067/HFDD6A30MA10 https://dx.doi.org/10.5067/M3ID6ZC0K2LW https://dx.doi.org/10.5067/VIIRS/VNP03IMG.001 https://dx.doi.org/10.5067/VIIRS/VNP03MOD.001 https://dx.doi.org/10.5067/NC505NHFIHCY https://dx.doi.org/10.5067/8ABZCV1TKE8D https://dx.doi.org/10.5067/9NPOTPIPLMAW https://dx.doi.org/10.5067/N9J1D8VZVJUX https://dx.doi.org/10.5067/VIIRS/VNP02DNB.001 https://dx.doi.org/10.5067/VIIRS/VNP02DNB_NRT.001 https://dx.doi.org/10.5067/VIIRS/VNP02GDC_NRT.001 https://dx.doi.org/10.5067/VIIRS/VNP02IMG.001 https://dx.doi.org/10.5067/VIIRS/VNP02MOD.001 https://dx.doi.org/10.5067/VIIRS/VNP02MOD_NRT.001 https://dx.doi.org/10.5067/VIIRS/VNP02OBC_NRT.001
Landsat 9 OLI-2 NASA, USGS	Approved (2020)	*Waveband categories:* SWIR, VIS, NIR *Feature/ phenomenon*	
Landsat 8 OLI USGS, NASA	Operational (2013)	*Waveband categories:* SWIR, VIS, NIR *Feature/ phenomenon*	
DMC3 VHRI-100 UKSA	Operational (2015)	*Waveband categories:* VIS, NIR *Feature/ phenomenon*	
SSTL S1-4 VHRI-100	Operational (2018)	*Waveband categories:* VIS, NIR	
UKSA		*Feature/ phenomenon*	

The pipeline output is a list of satellites, instruments, and datasets, listed in order according to how well they match the initial need criteria. We have further improvements planned for this pipeline, including a user interface, and will announce when it is ready for public use!

Acknowledgment

This research is supported by NASA Grant NNM11AA01A as part of the Interagency Implementation and Advanced Concepts Team (IMPACT) program.