Machine Learning for NASA Earth Science Data Systems

Manil Maskey, Ph.D.
Lead, NASA SMD Strategic Data Management Working Group on AI/ML, NASA HQ
Senior Research Scientist, Earth Science Data Systems, NASA HQ
Deputy Manager, NASA IMPACT, MSFC

The 2nd NOAA Workshop on Leveraging AI in Environmental Science
February 11, 2021

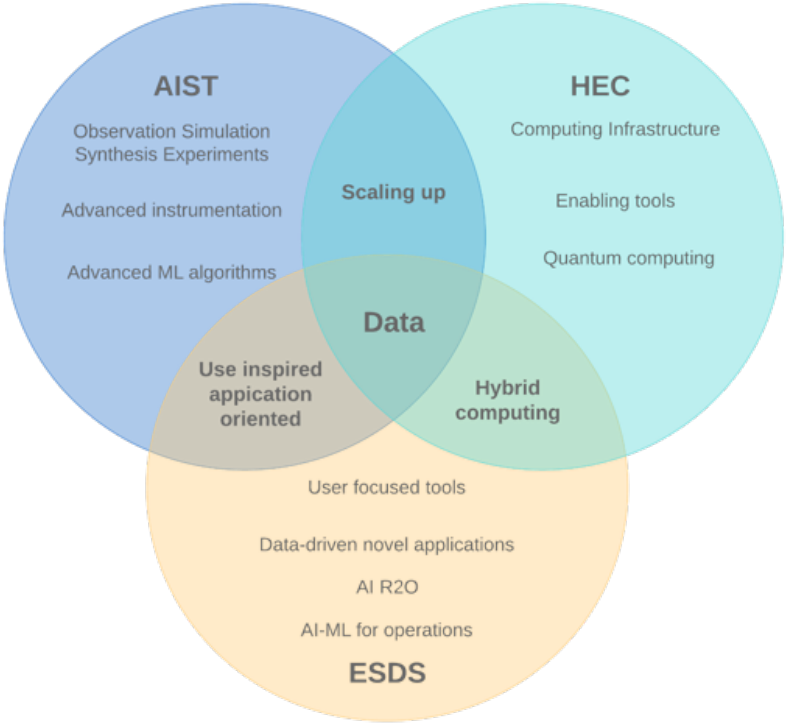
NASA Science Mission Directorate (SMD) Artificial Intelligence and Machine Learning (AI/ML) Initiatives

SMD's Strategy for Data Management and Computing for Groundbreaking Science 2019-2024 Report identified that AI/ML has yet to be fully appreciated and understood by SMD and science disciplines

Activities:

- Identify areas of natural collaborations on AI/ML across SMD
- Conduct expert workshop on AI for science
- Explore industry partnership to transform data systems and leverage open science data
- Develop a roadmap to leverage large volumes of data and computation to accelerate AI/ML across NASA science

ESD Programs Supporting AI/ML



R&A – individual PIs may use AI-ML techniques tailored to disciplines

ESD has invested in AI-ML across the Division

- Algorithms and Hardware infrastructure
- Data is key element of AI systems
- Cloud Computing
- Catalogs and Standardization

ESDS AI/ML strategic goals

Goal 1: Augment scientific data stewardship processes

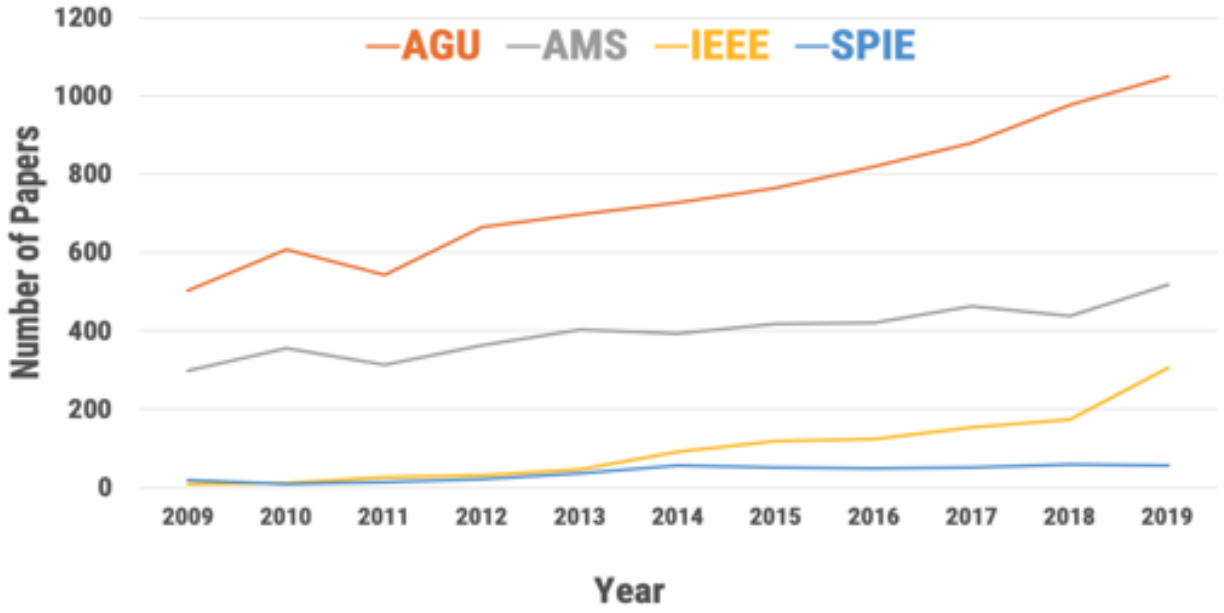
Goal 2: Maximize information and knowledge discovery capabilities

Goal 3: Enable sharing, and interoperability of Earth science AI training data and models

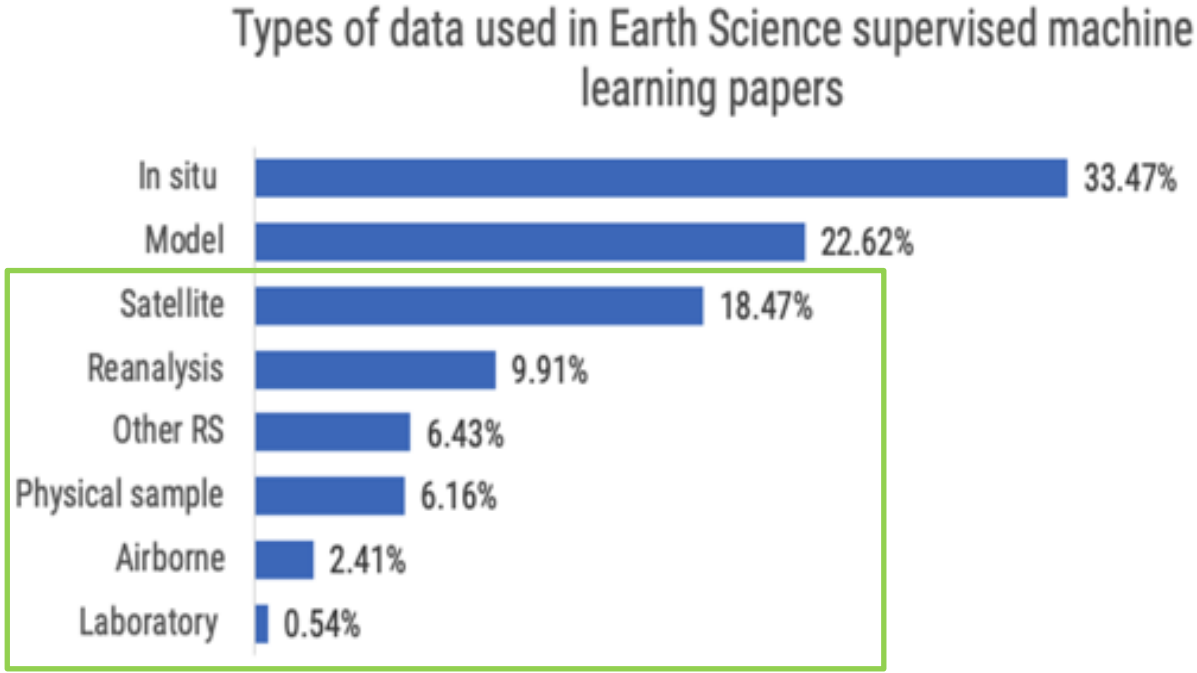
Goal 4: Increase engagement with commercial, academic, other agencies, and international partners

Goal 5: Foster AI expertise within the program

AI/ML in Earth Science

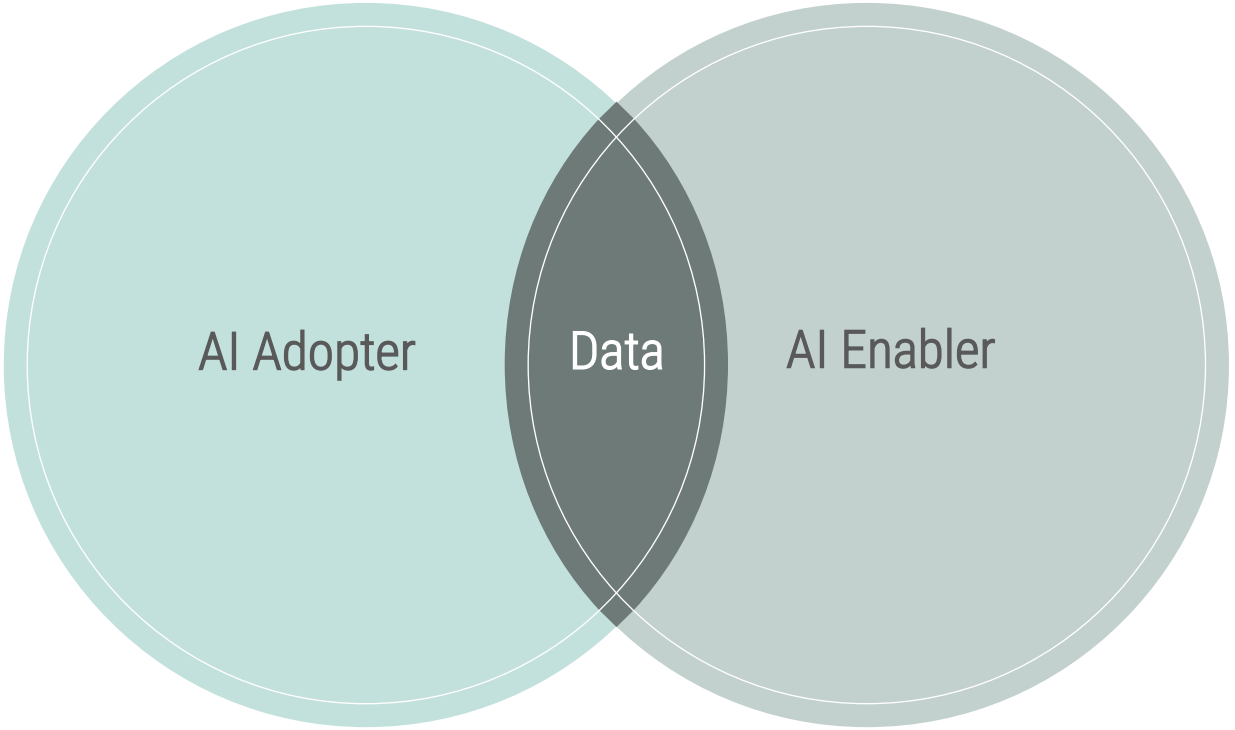


Rapid adoption of AI/ML by Earth science researchers



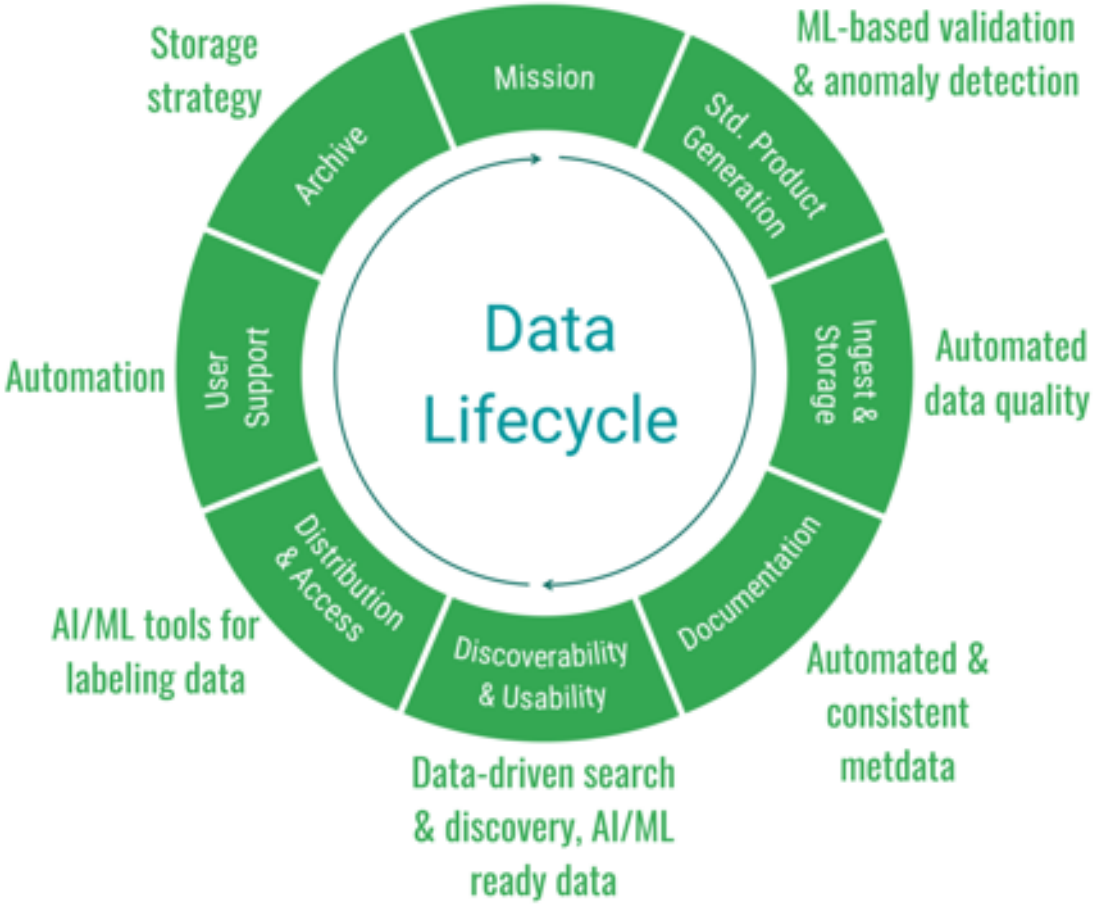
ESDS archive is underutilized by AI applications, and new data curation activities are needed to address the lack of training data

AI and Data Systems



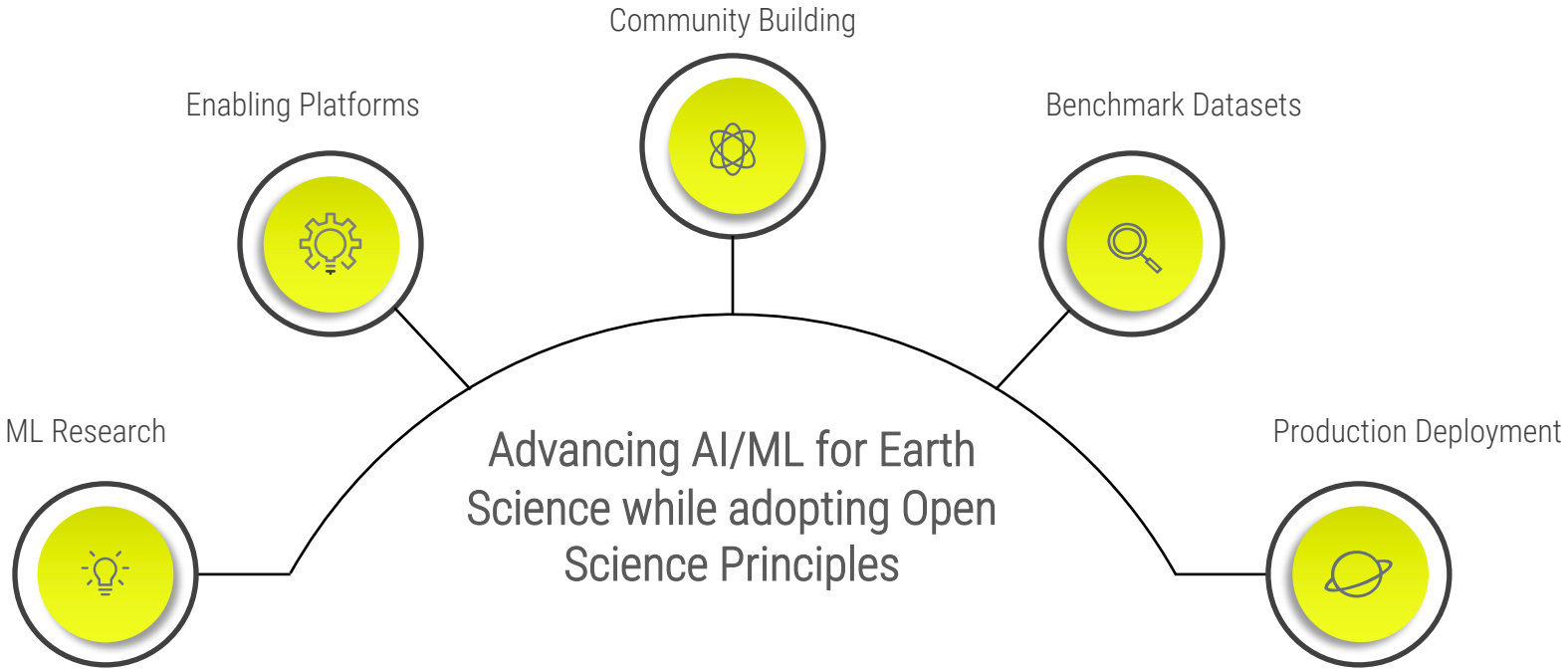
Data systems as adopter of AI

- Improve efficiency of NASA's data systems operations
- Increase opportunity for researchers and commercial users to access/process PBs of data quickly without the need for data management
- Ensure users find right data for their problem
- Minimize user burden to access data
- Enable users to extract new knowledge/information from archives

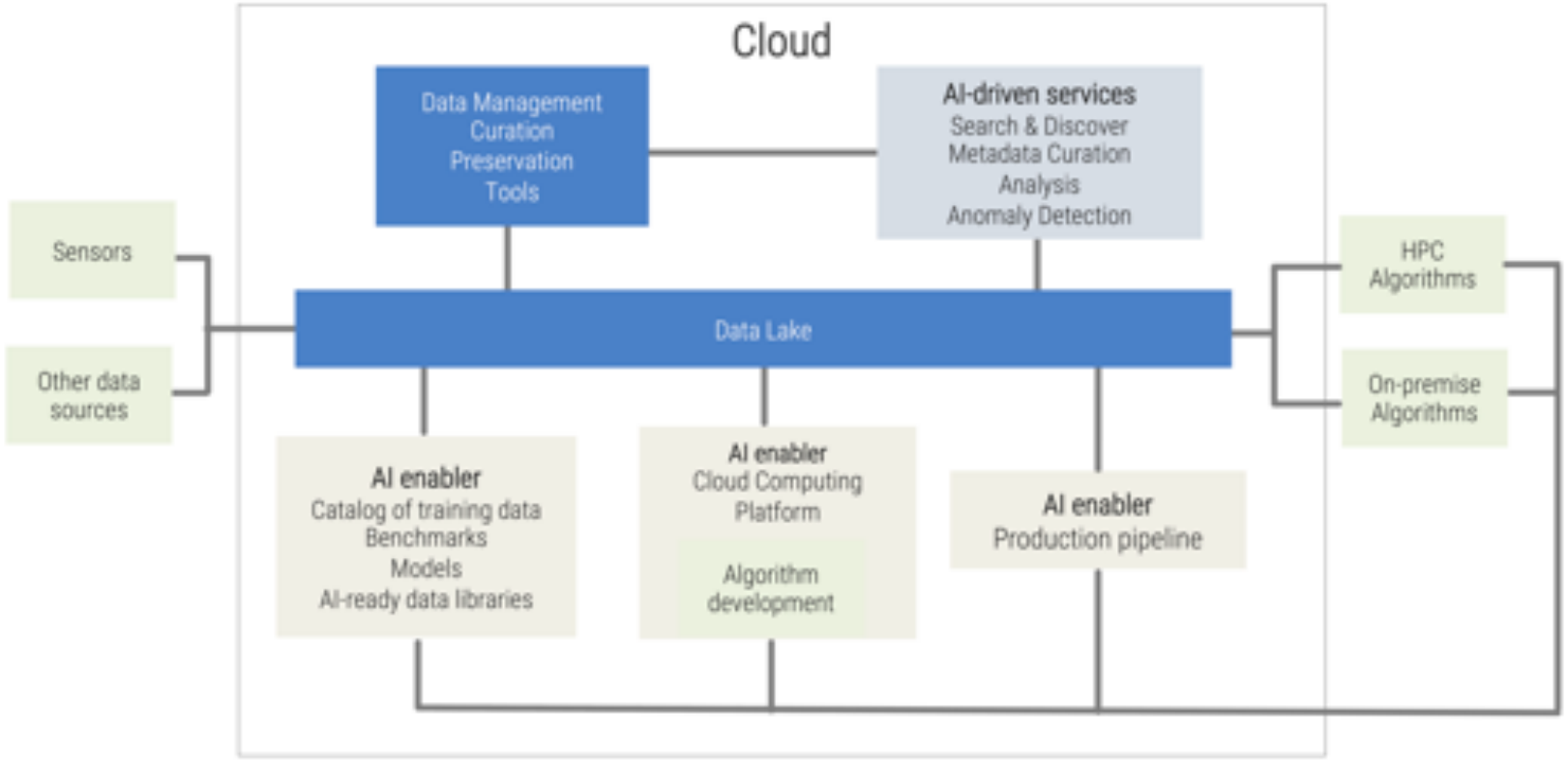


AI for full data lifecycle

Data systems as enabler of AI



Data systems AI architecture design



Current projects/activities

Maximize information and knowledge discovery capabilities – computer vision
Novel search of archived data

Augment data stewardship processes – NLP/Knowledge Graph
Keyword assignment
Metadata curation

Accelerate AI for Earth science – AI-ready data
Labeled dataset generation/stewardship
Data science challenges
Labeling tools

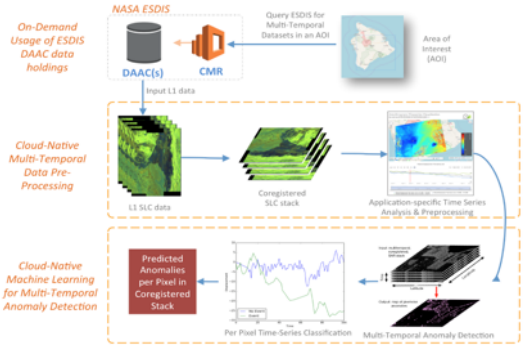
Enabling Platforms
Cloud/HPC
Deployment pipeline

Partnerships
Industry/Academia involvement
Workshops/Hackathons

Ongoing AI/ML activities

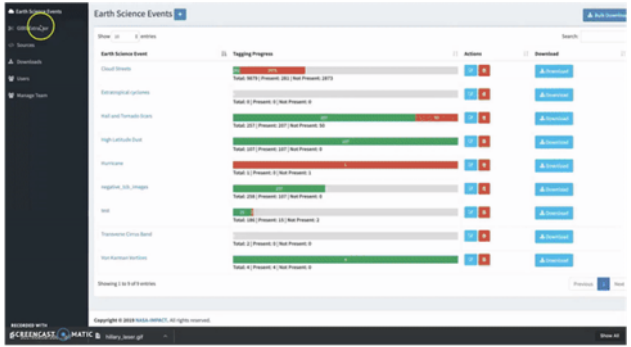
ACCESS

2017: Multi-temporal anomaly detection for SAR
2019: Several AI/ML focused projects

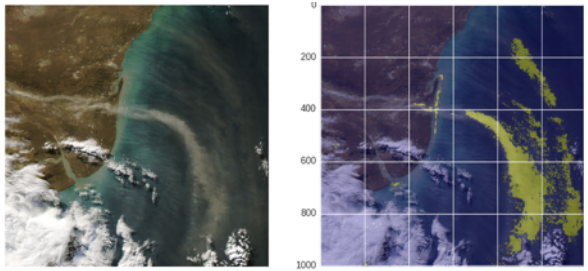


Workshops

Labeling tool

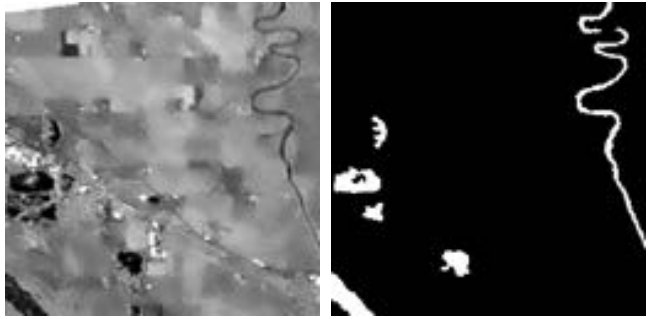


Rapid prototypes



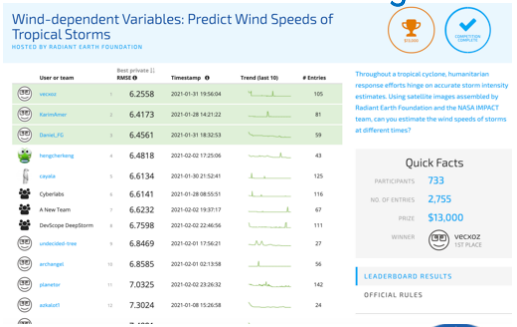
High latitude dust detection

Training dataset generation

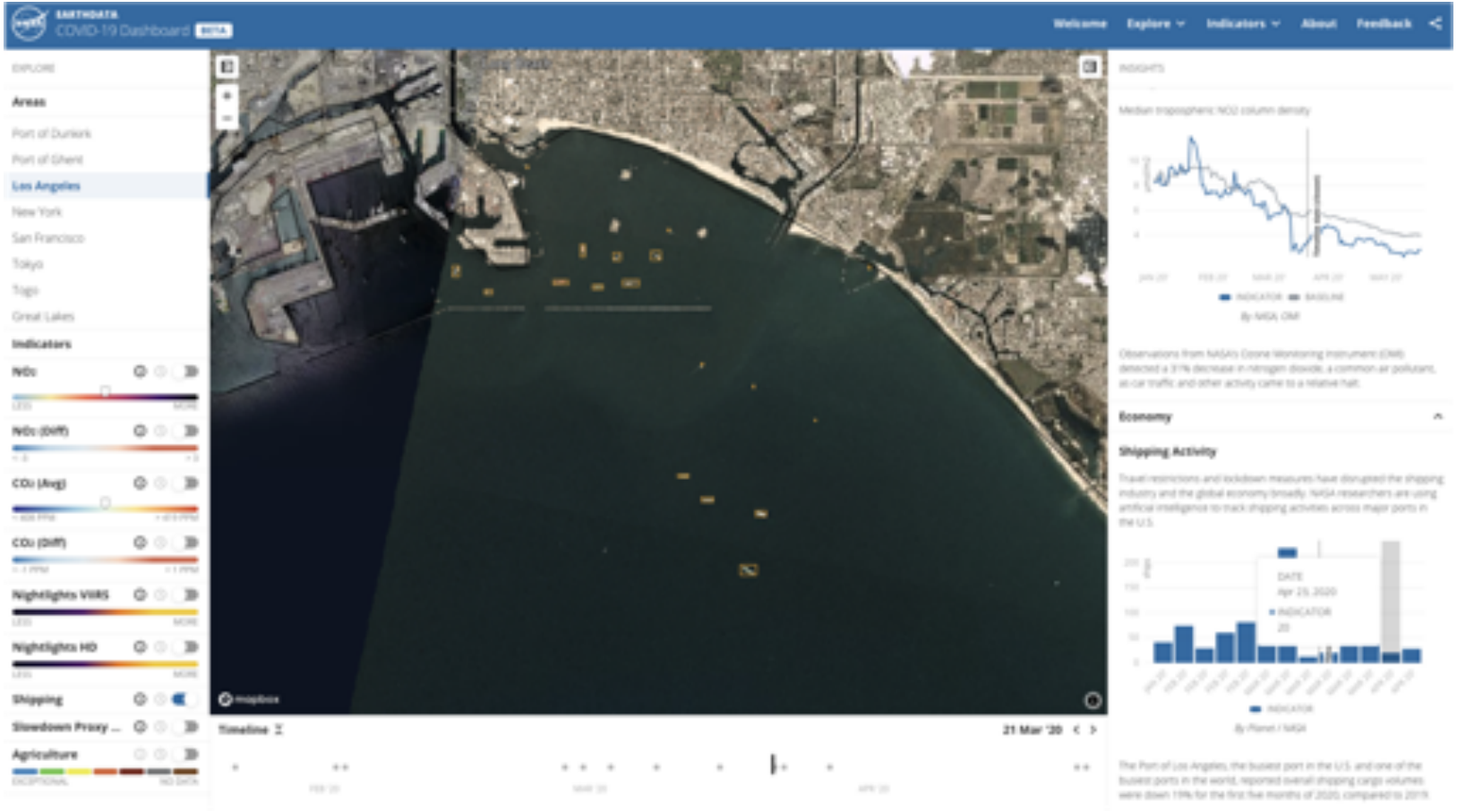


Flood extent dataset on Sentinel 1

Data science challenges



Leveraging AI+CSDAP+Prototypes+Cloud for COVID-19 Dashboard



Resources

Image labeler	https://labeler.nasa-impact.net/
Phenomena portal	http://phenomena.surge.sh/
Hurricane intensity estimation portal	http://hurricane.dsig.net/
COVID-19 dashboard	https://earthdata.nasa.gov/covid19/
IMPACT website	https://impact.earthdata.nasa.gov/
ACCESS 2019 projects	https://earthdata.nasa.gov/esds/competitive-programs/access

Thank you.

Manil Maskey
manil.maskey@nasa.gov

Backup slides

Maximize information and knowledge discovery capabilities

Why?

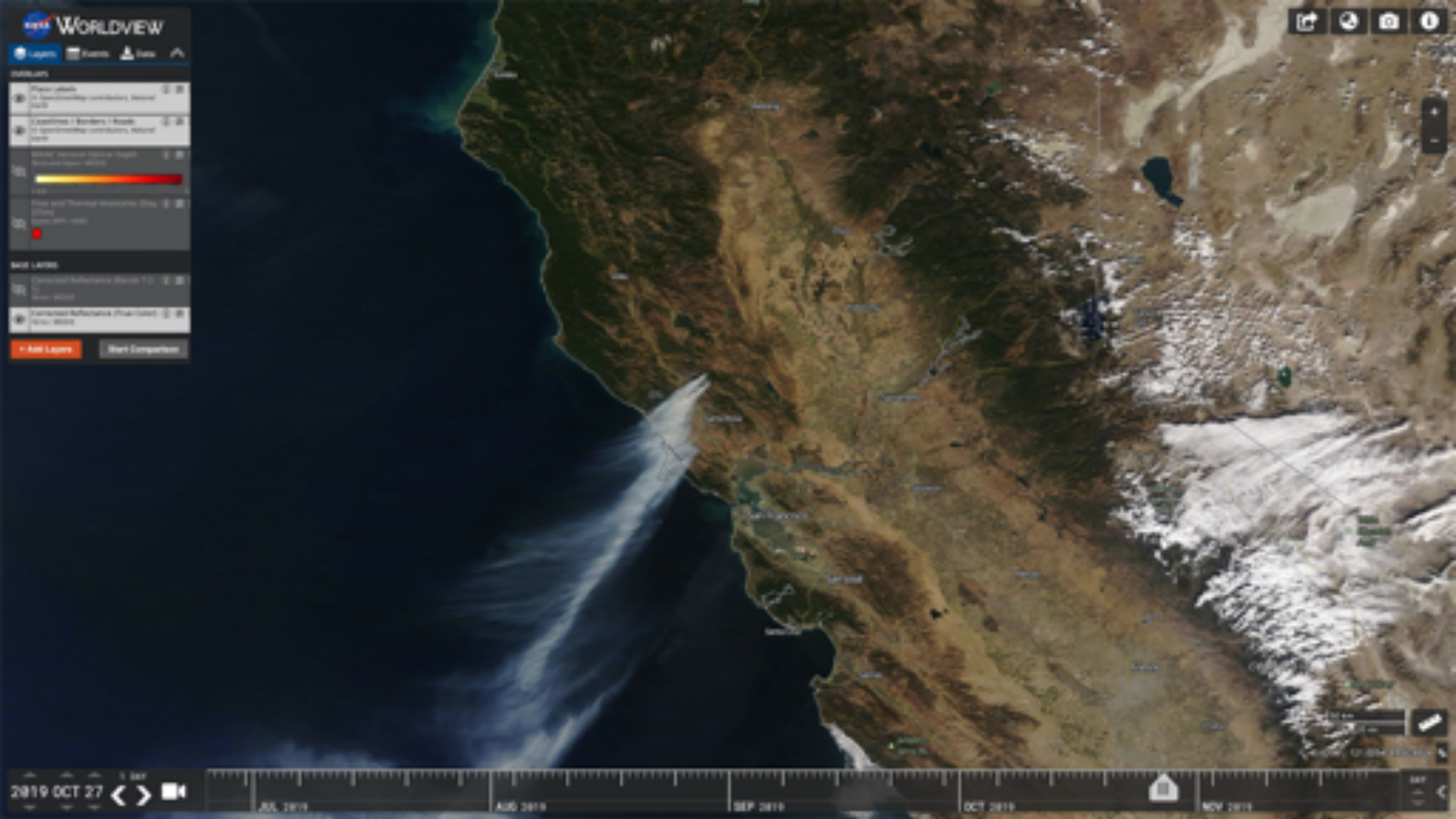
Increasing Earth science data archives require non-traditional approaches to data management

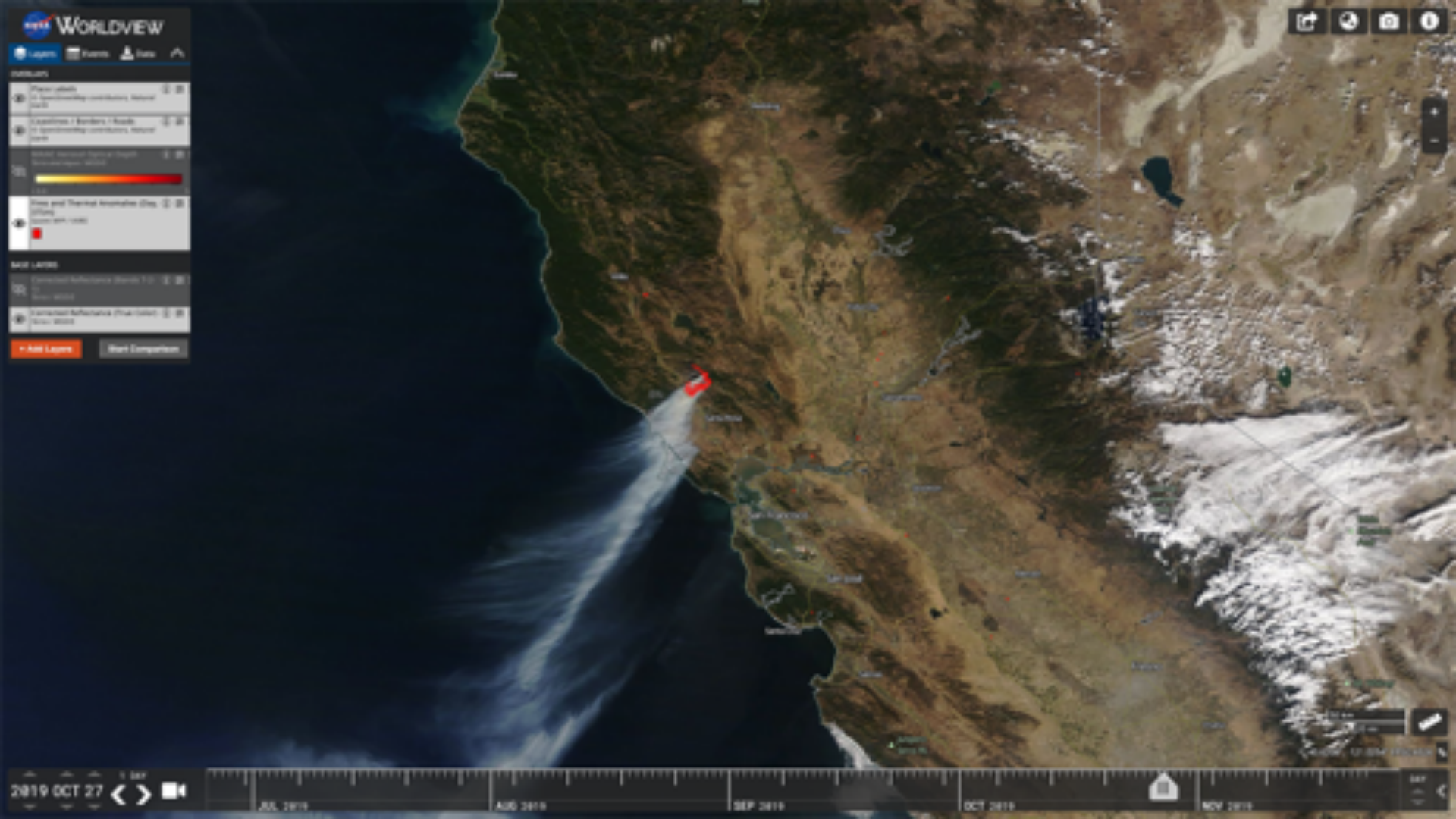
Data driven technologies to provide advanced search capabilities

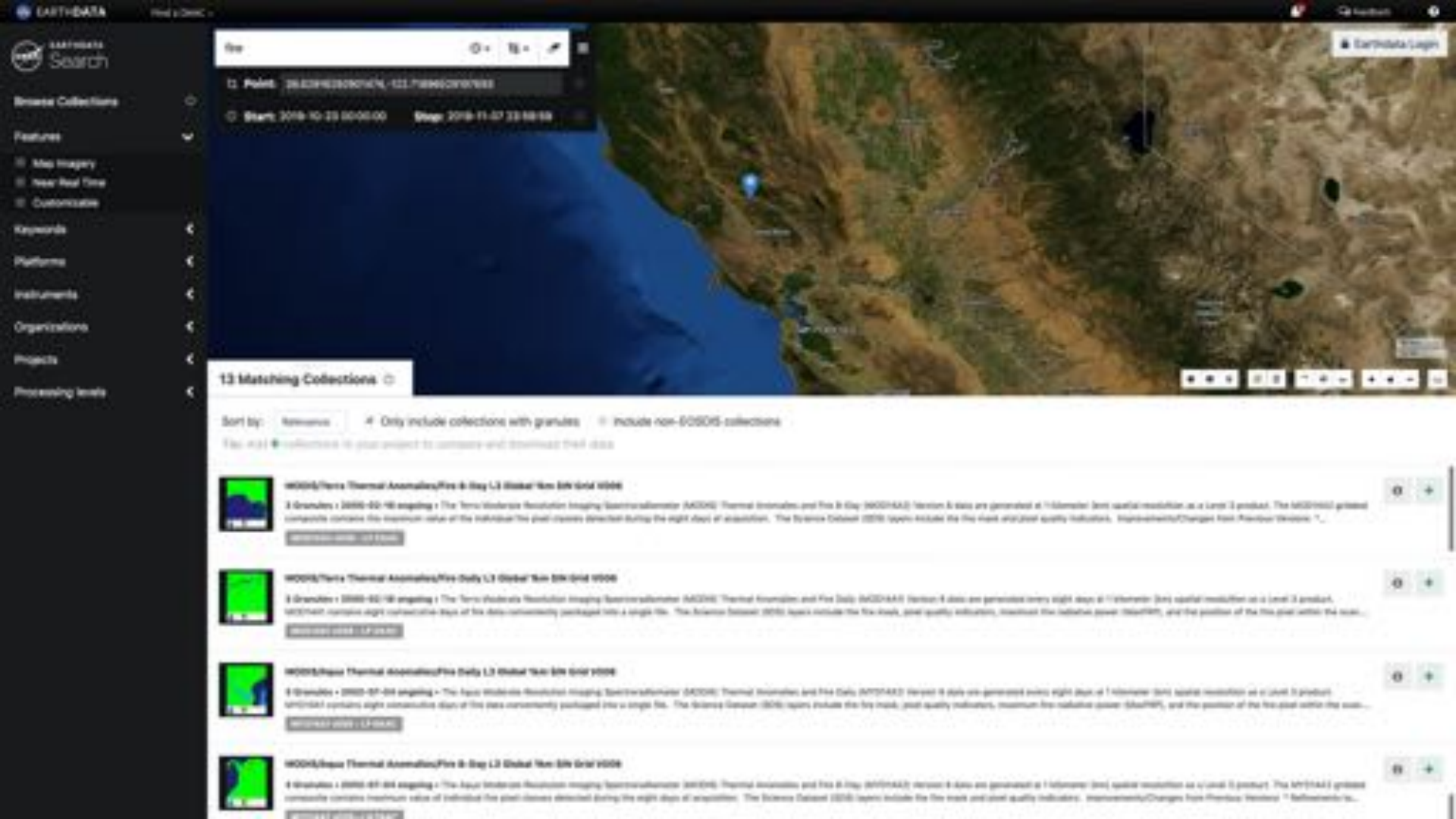
Machine learning-based approach - an enabling data driven technology to provide automated detection of Earth science events from image archives

Catalog of events can provide a novel way to explore large archives of data

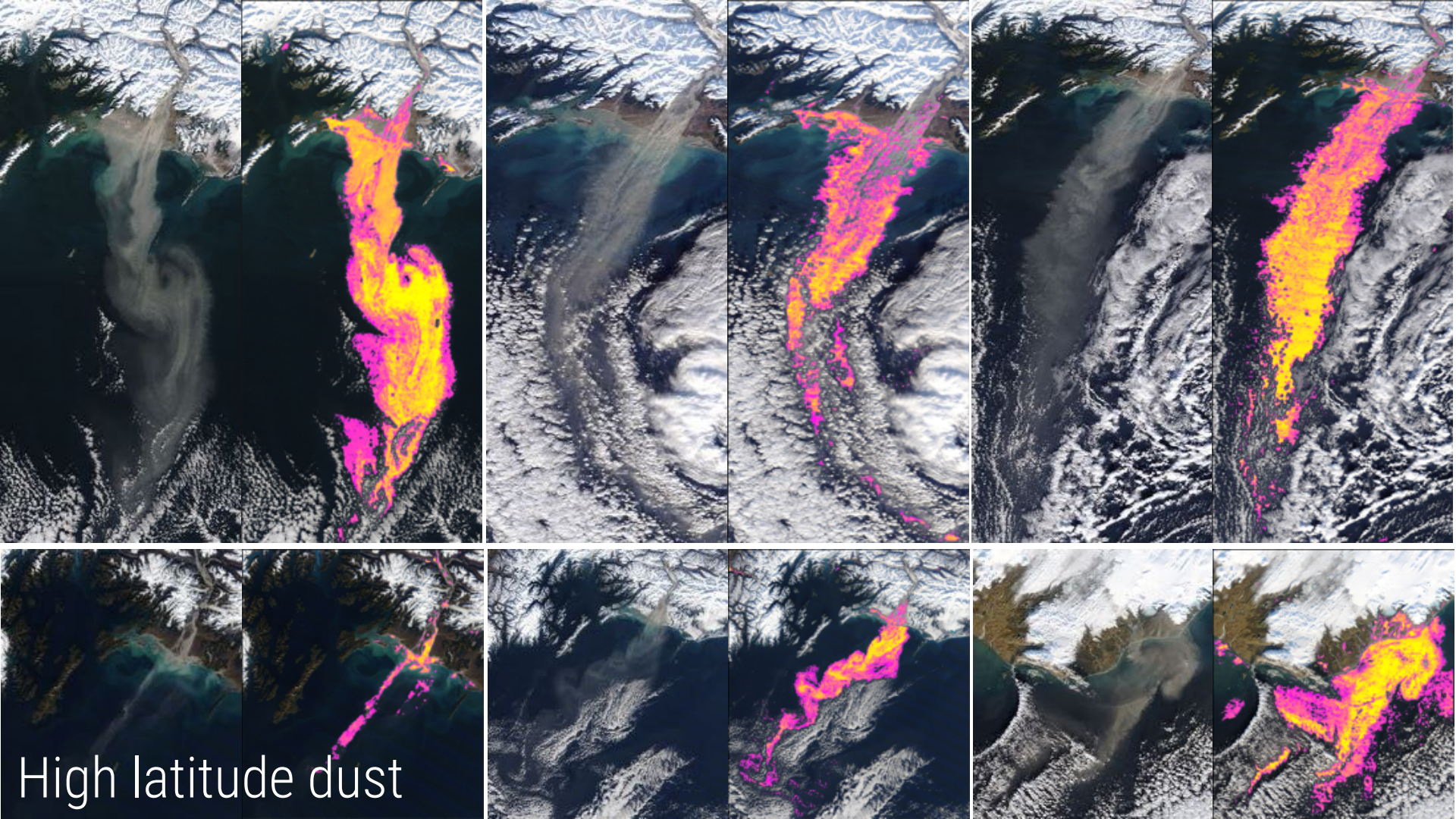
Discover and explore Earth science data archives around events using machine learning (ML) techniques


WORLDVIEW
Layers
Events
Data
OVERLAYS
BASE LAYERS
+ Add Layers
Start Comparison
2019 OCT 27
1 DAY
JUL 2019
AUG 2019
SEP 2019
OCT 2019
NOV 2019

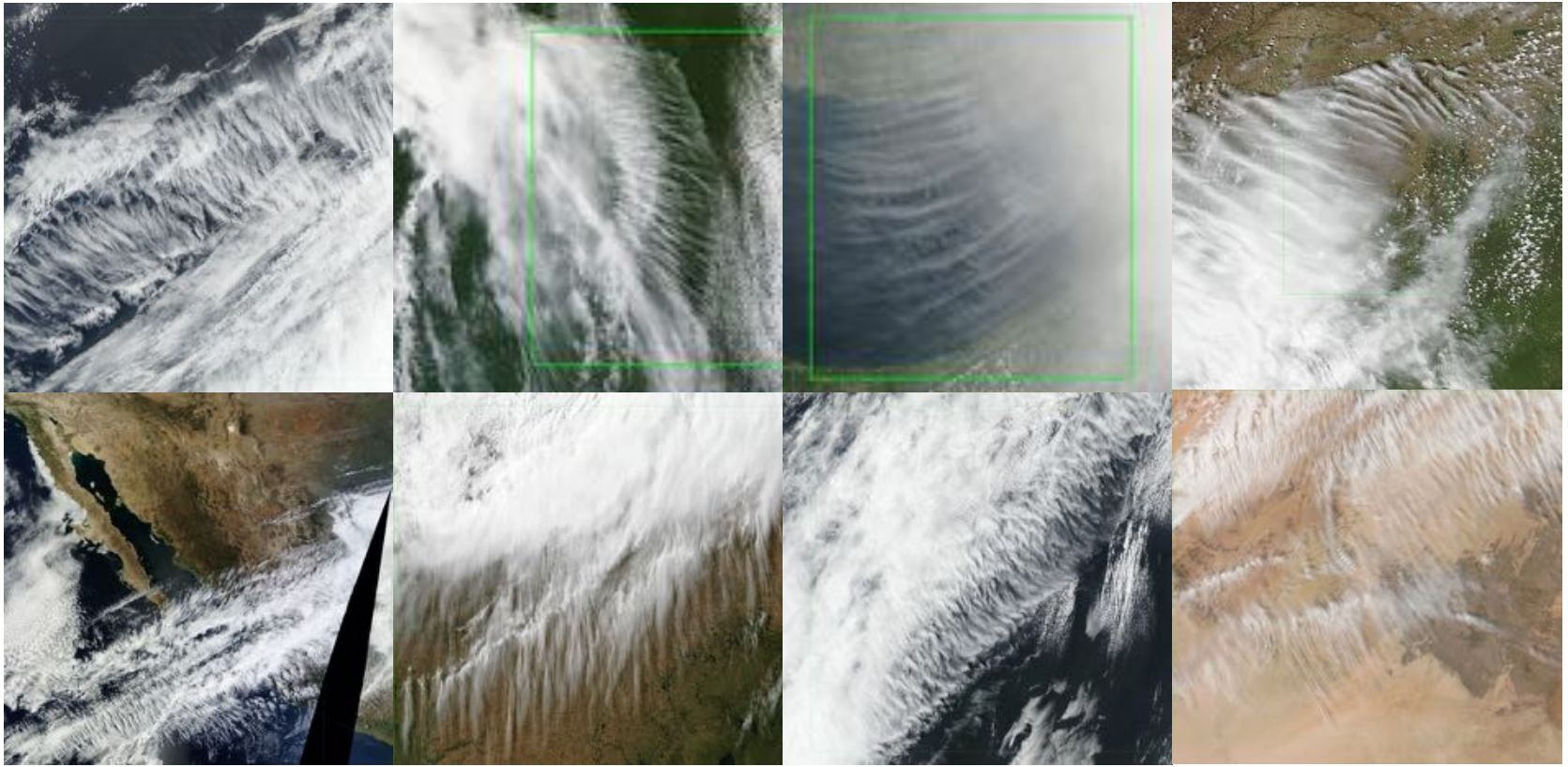

WORLDVIEW
Layers
Events
Data
OVERLAYS
BASE LAYERS
+ Add Layers
Start Comparison
2019 OCT 27
1 DAY
JUL 2019
AUG 2019
SEP 2019
OCT 2019
NOV 2019

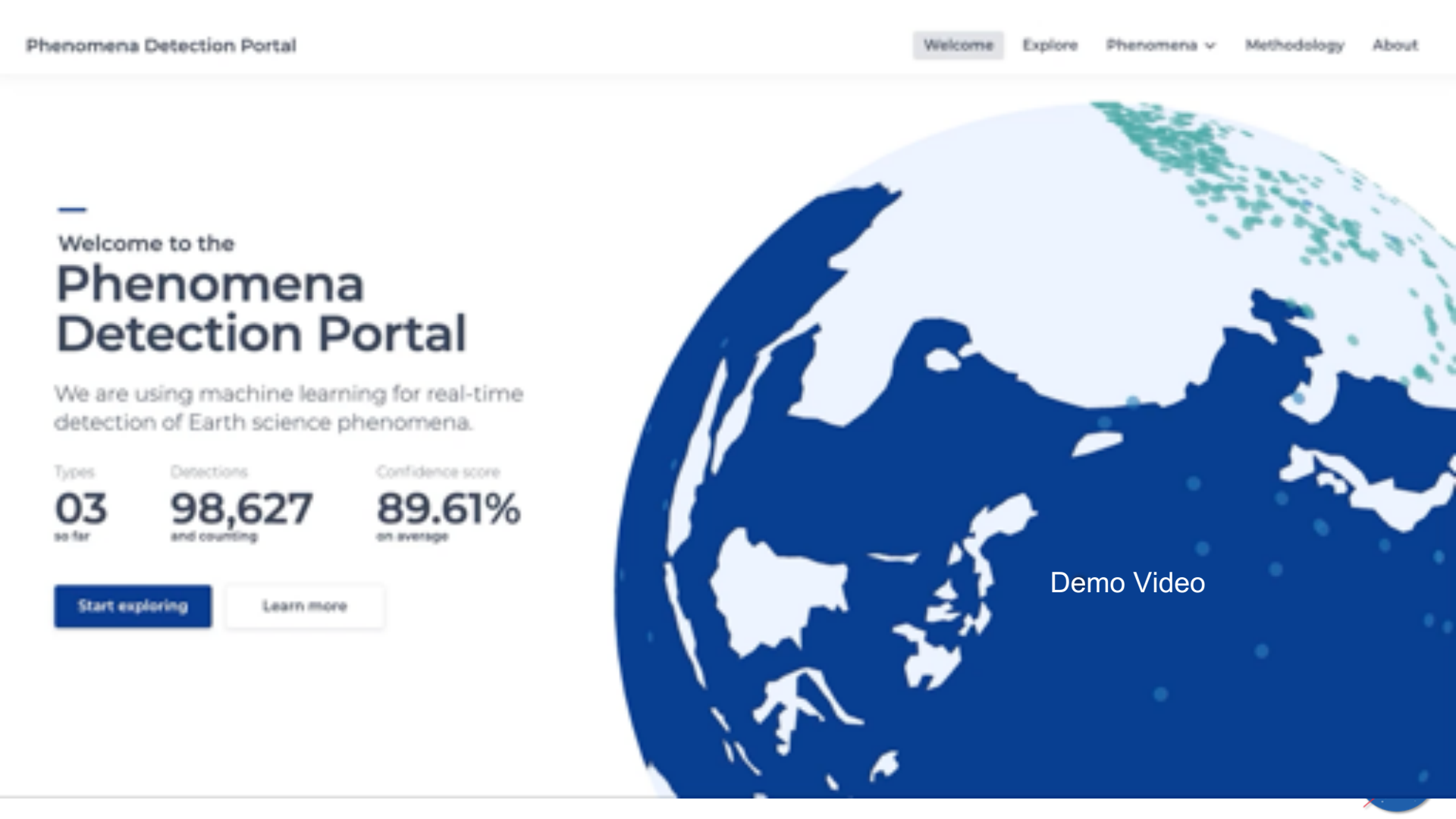

EARTHDATA
Search
Browse Collections
Features
Map Imagery
Near Real Time
Customizable
Keywords
Platforms
Instruments
Organizations
Projects
Processing levels
fire
Point:
Start: 2019-10-23 00:00:00
Stop: 2019-11-07 23:59:59
Earthdata Login
13 Matching Collections
Sort by:
Only include collections with granules
Include non-EOSDIS collections


Demo Video


High latitude dust

Transverse cirrus bands




Phenomena Detection Portal
Welcome
Explore
Phenomena
Methodology
About
Welcome to the
Phenomena Detection Portal
We are using machine learning for real-time detection of Earth science phenomena.
Types
03
so far
Detections
98,627
and counting
Confidence score
89.61%
on average
Start exploring
Learn more
Demo Video

Augment data stewardship processes

Automated keyword assignment

Why?

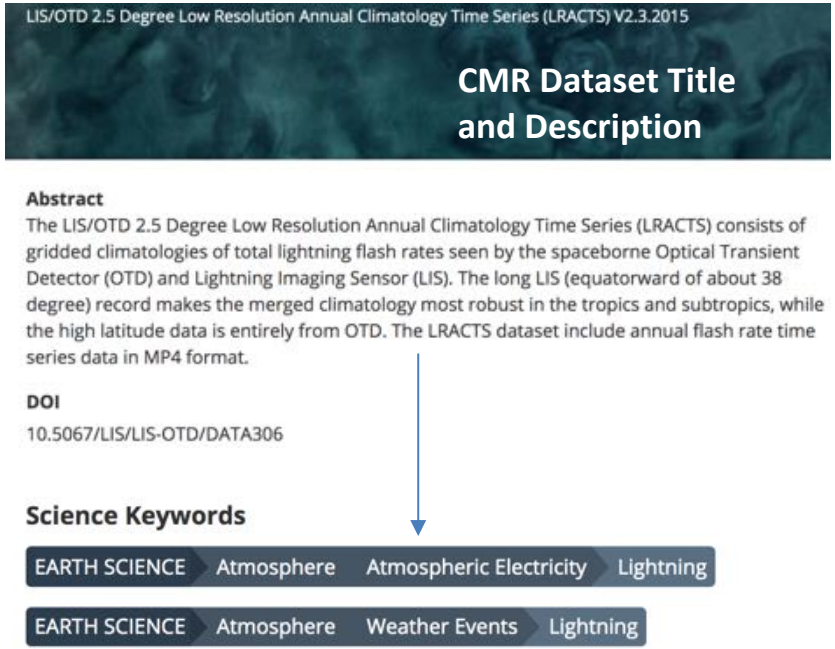
Assigning science keywords is currently a manual process, which is prone to human error and inconsistencies.

Metadata managed across a network of multiple data centers (i.e. keywords not assigned by a central entity)

Keywords may be assigned by non-subject matter experts (SMEs)

Improve metadata quality

Provide objective and consistent approach to keyword assignment



Approach – build word embeddings

88,410

documents

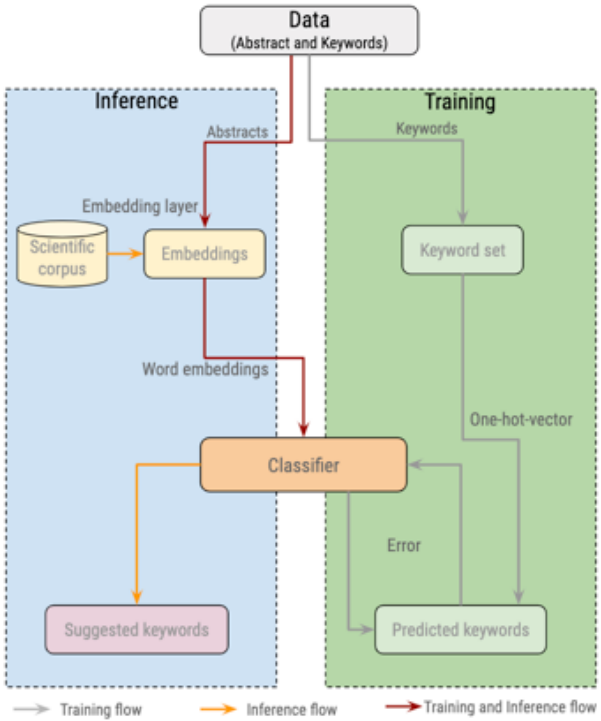
530 million

words

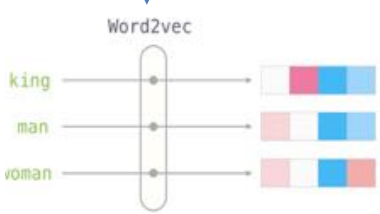
5.5 million

unique words

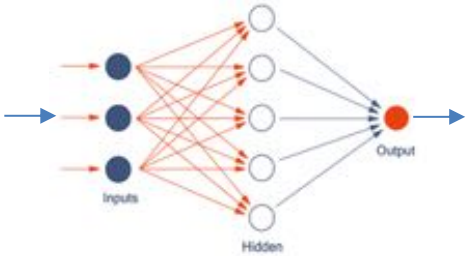
Automated keyword assignment



Version 7.3 is the current version of the data set. Version 3.5 is no longer available and has been superseded by Version 7.3. This data set is currently provided by the OCO (Orbiting Carbon Observatory) Project. In expectation of the OCO-2 launch, the algorithm was developed by the Atmospheric CO2 Observations from Space (ACOS) Task as a preparatory project, using GOSAT TANSO-FTS spectra. After the OCO-2 launch, "ACOS" data are still produced and improved, using approaches applied to the OCO-2 spectra. The "ACOS" data set contains Carbon Dioxide (CO2) column averaged dry air mole fraction for all soundings for which retrieval was attempted. These are the highest-level products made available by the OCO Project, using TANSO-FTS spectral radiances, and algorithm build version 7.3. The GOSAT team at JAXA produces GOSAT TANSO-FTS Level 1B

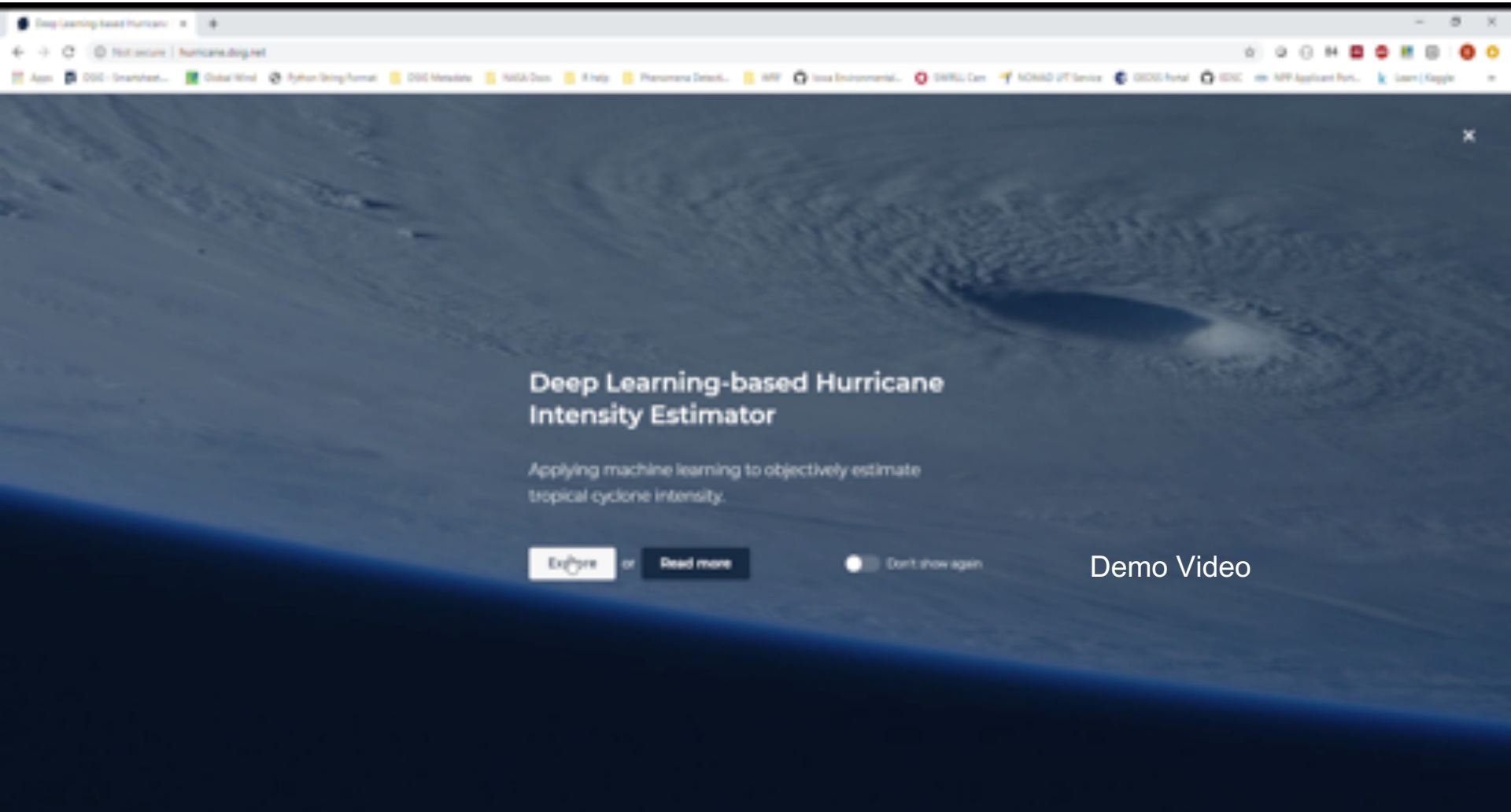


word2vec



classifier

Predicted Keyword	Score
carbon dioxide	0.4513424
land use/land cover classification	0.3825603
terrain elevation	0.1924277
barometric altitude	0.18085223
carbon and hydrocarbon compounds	0.07634798


Deep Learning-based Hurricane Intensity Estimator
Applying machine learning to objectively estimate tropical cyclone intensity.
Explore
or
Read more
Don't show again
Demo Video