NASA's EO and ML Road Map

Dr. Rahul Ramachandran

Manager | Inter-Agency Implementation and Advanced Concepts Team (IMPACT)
Senior Research Scientist | Earth Science Branch (ST11)
Marshall Space Flight Center/NASA
Huntsville, Alabama 35812, USA

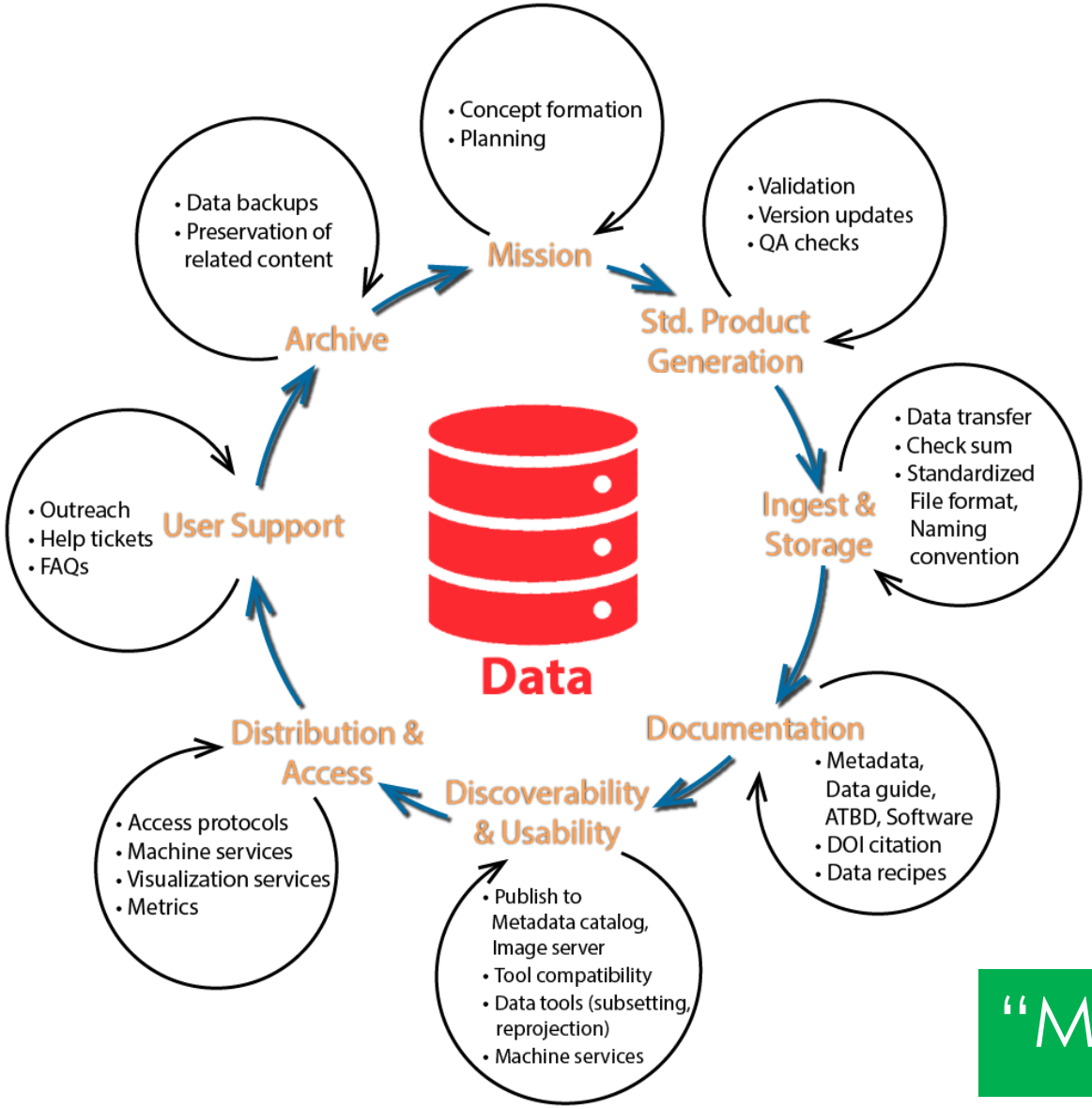
Trillion Pixel Challenge, Oak Ridge

Earth Science – NASA's Strategic Goal

This ability to ***observe our planet comprehensively*** matters to each of us, on a daily level. Earth information - for use in Internet maps, daily weather forecasts, land use planning, transportation efficiency, and agricultural productivity, to name a few - is central to our lives, providing substantial contributions to our economies, our national security, and our personal safety. It helps ensure we are a thriving society. - NRC, 2018

NASA's Strategic Goal 1.1: "Understand The Sun, Earth, Solar System, And Universe."

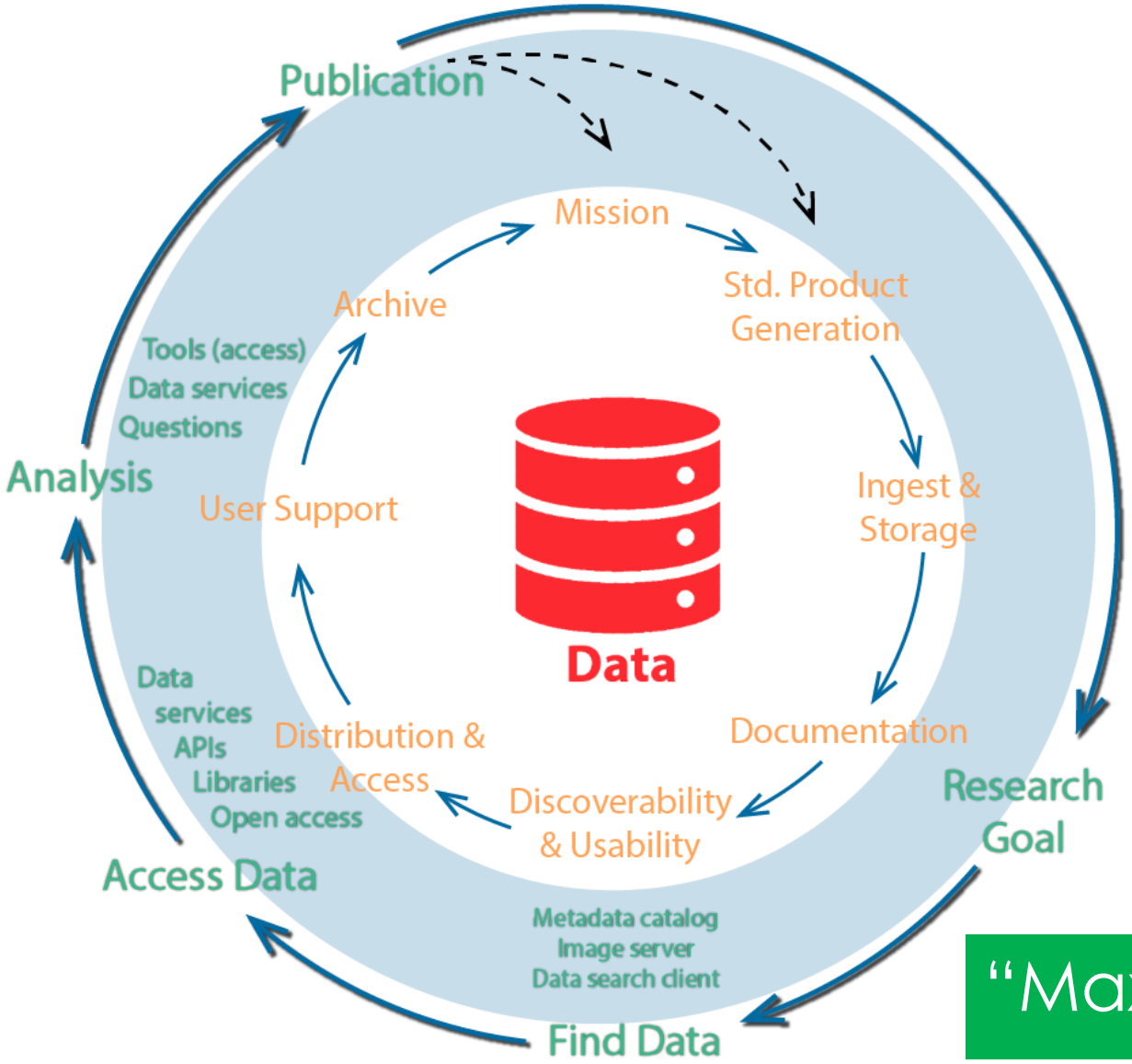
Earth Science Data System Program



- Actively manage NASA's Earth science data (Satellite, Airborne, and Field) through the entire data life cycle
- Use systematic and engineered process for data stewardship
- Process (and reprocess) instrument data to create high quality long-term Earth science data records
- Uphold NASA's policy of full and open sharing of all data, tools, and ancillary information for all users

"Maximize return on NASA data sets"

Earth Science Data System Program



- Develop and provide unique data system capabilities optimized to support rigorous science investigations and interdisciplinary research
- *Data system supports the research life cycle*
- Utilize standards to ensure interoperability with other systems/agencies/organization
- Engage members of the Earth science community to evolve the data systems

"Maximize return on NASA data sets"

NASA's Earth Science Data System in 2018

EOSDIS currently has over **27 Petabytes** of accessible Earth science data

Easy access and discovery of data to over **12,500 unique data products**

… of which 95% of granule searches complete in less than **1 Second**

EOSDIS delivered over **1.6 Billion** data products to over **4.1 Million** users from around the world

33,000 Data Collections in the Common Metadata Repository (CMR)

EOSDIS also delivers near-real-time products in under **3 hours** from observation …

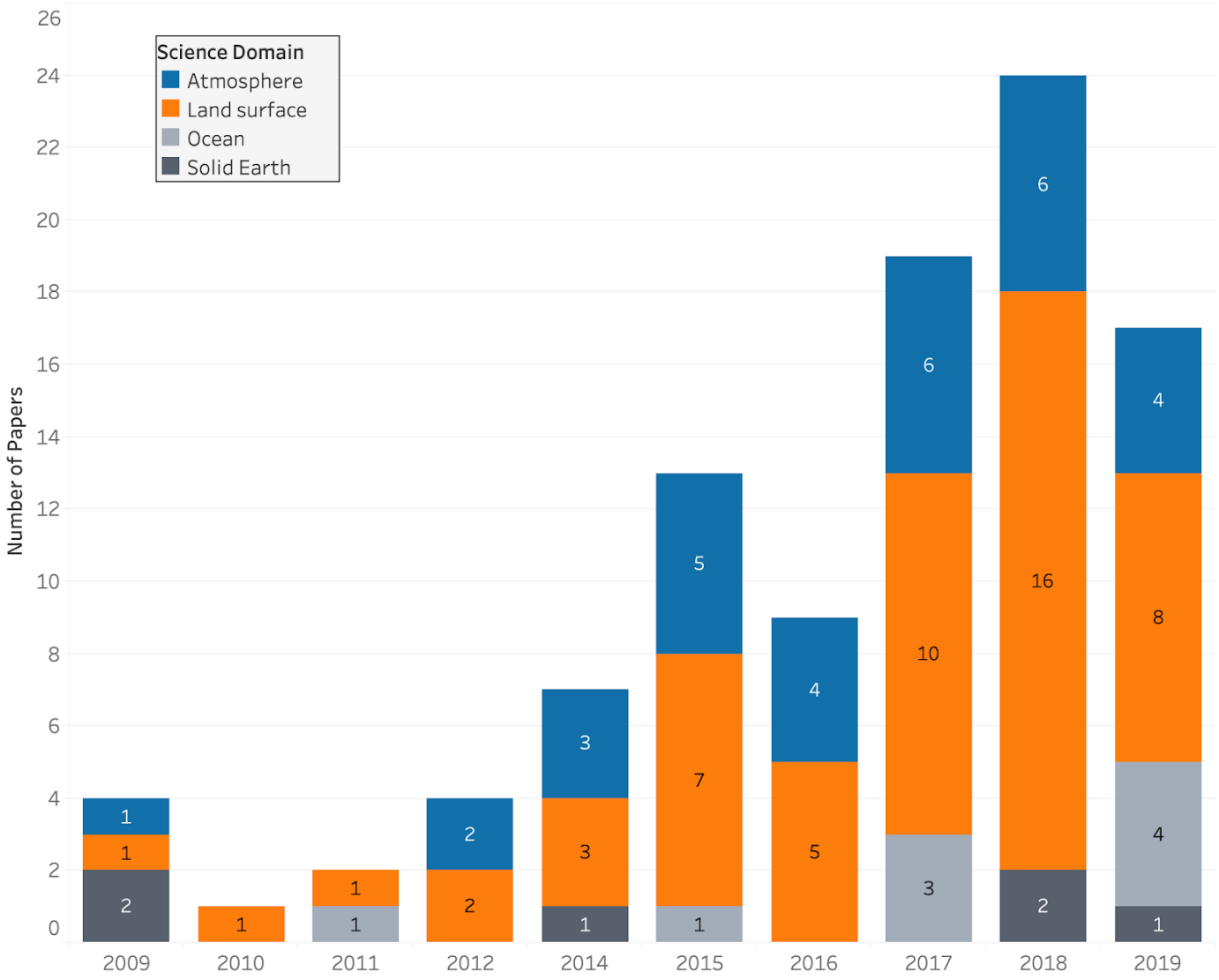
Over **330,000 users** have registered with EOSDIS to date

And Over **380 Million** data granules

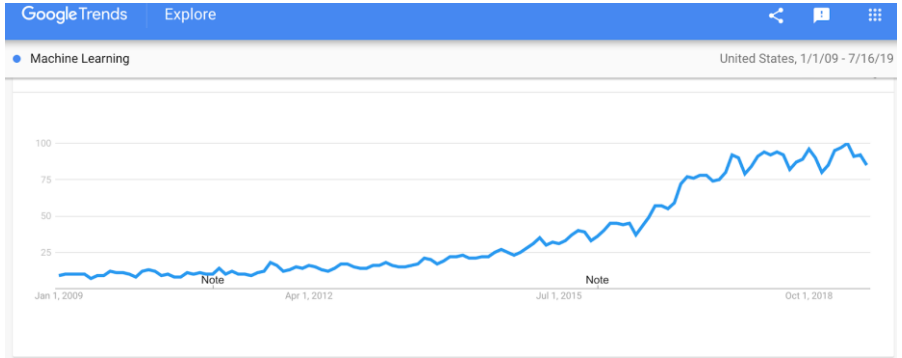
American Customer Satisfaction Index (ACSI) survey scoring **79** from over **4,000** respondents

https://earthdata.nasa.gov/about/system-performance

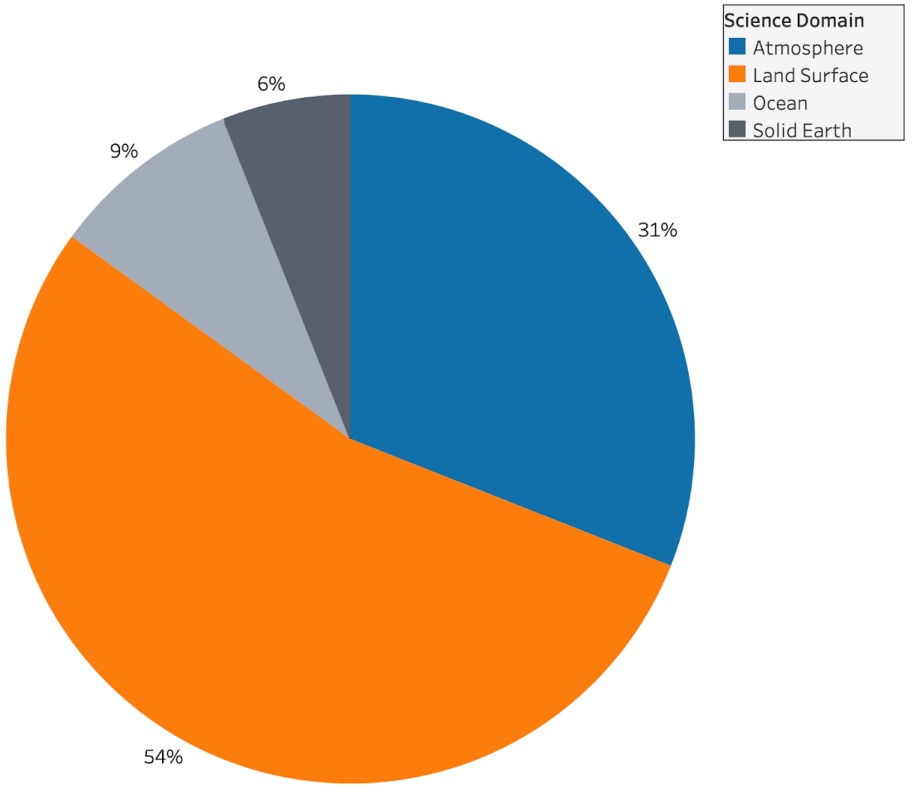
ML in Earth Science: Current Landscape



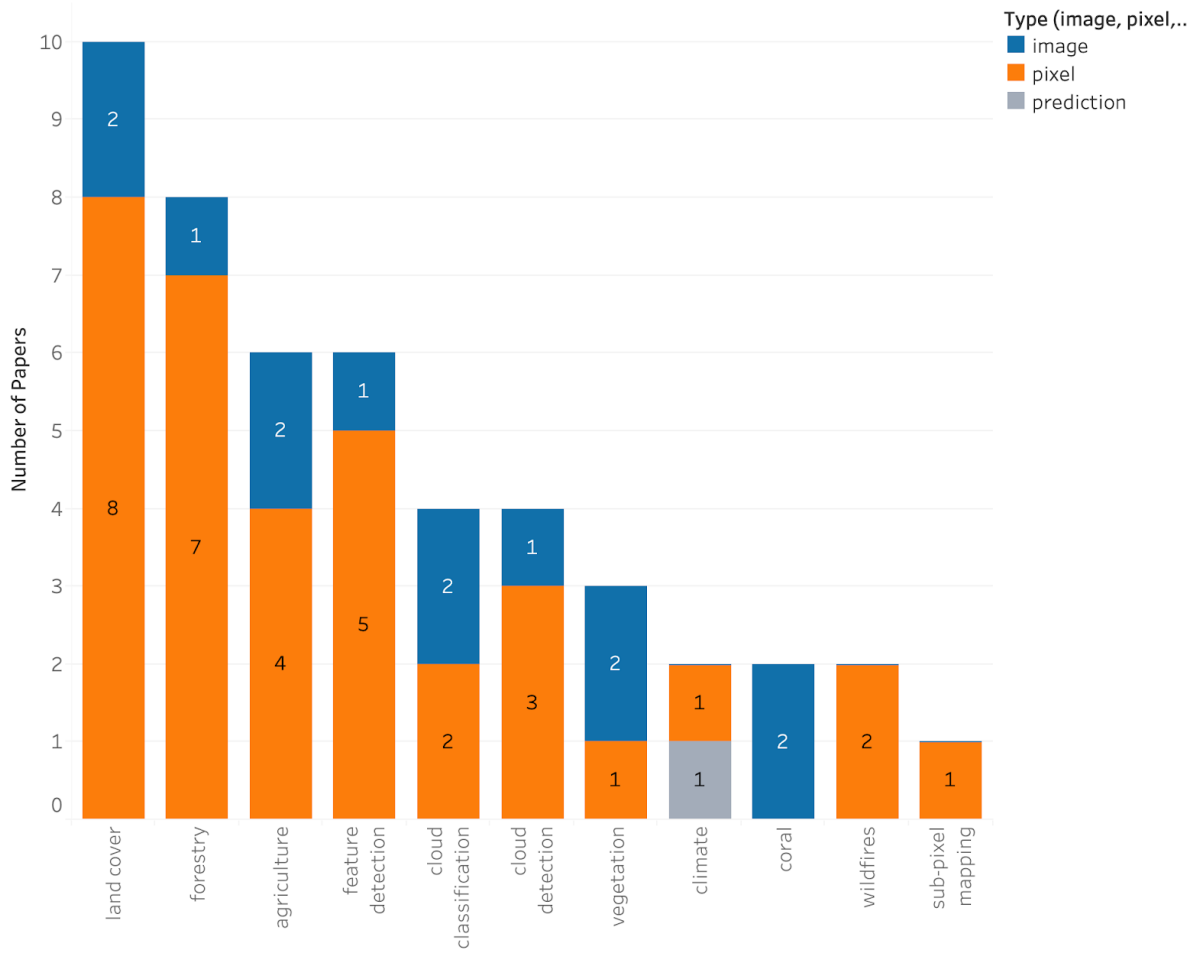
- Survey of ML application papers in all Earth Science journals from IEEE, SPIE, Elsevier, AMS, AGU (98 papers met search criteria)
- Time period: 2009-2019 (~10 years)



ML in Earth Science: Current Landscape

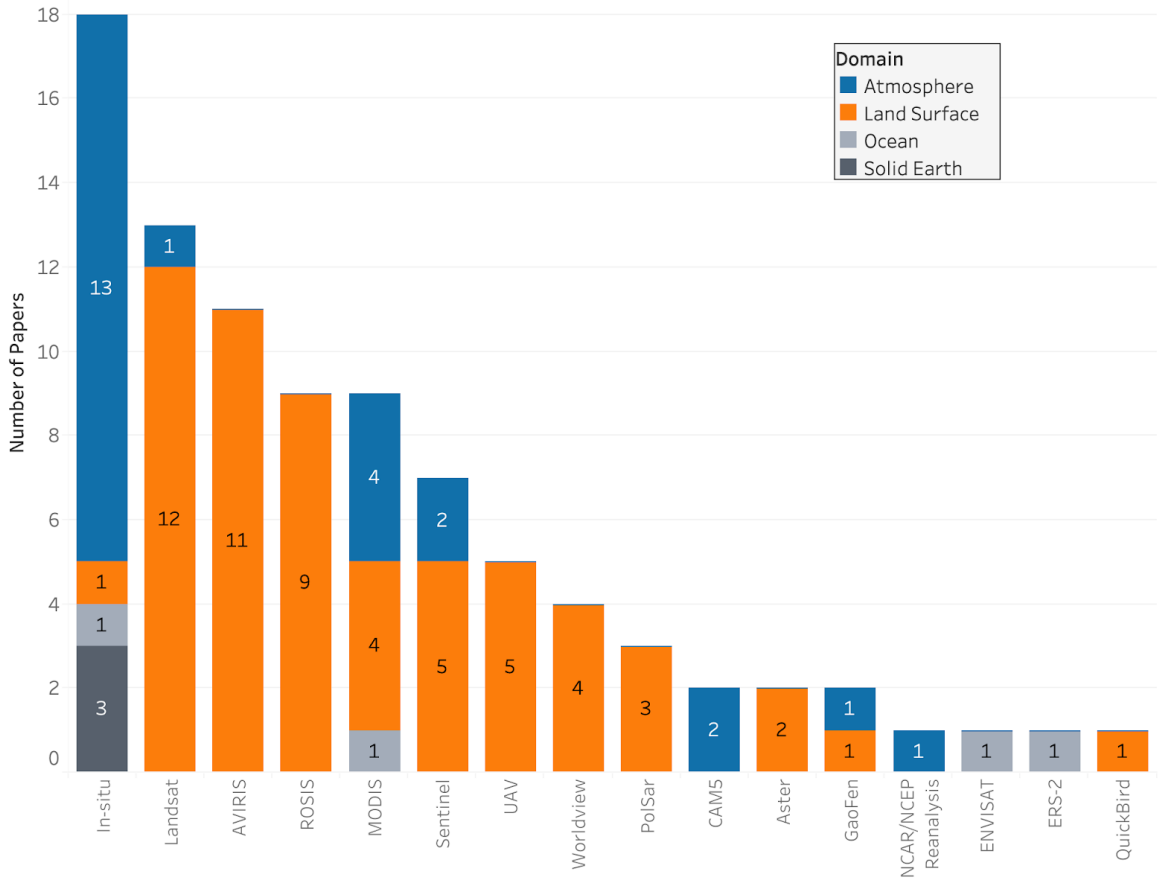


Science Domain

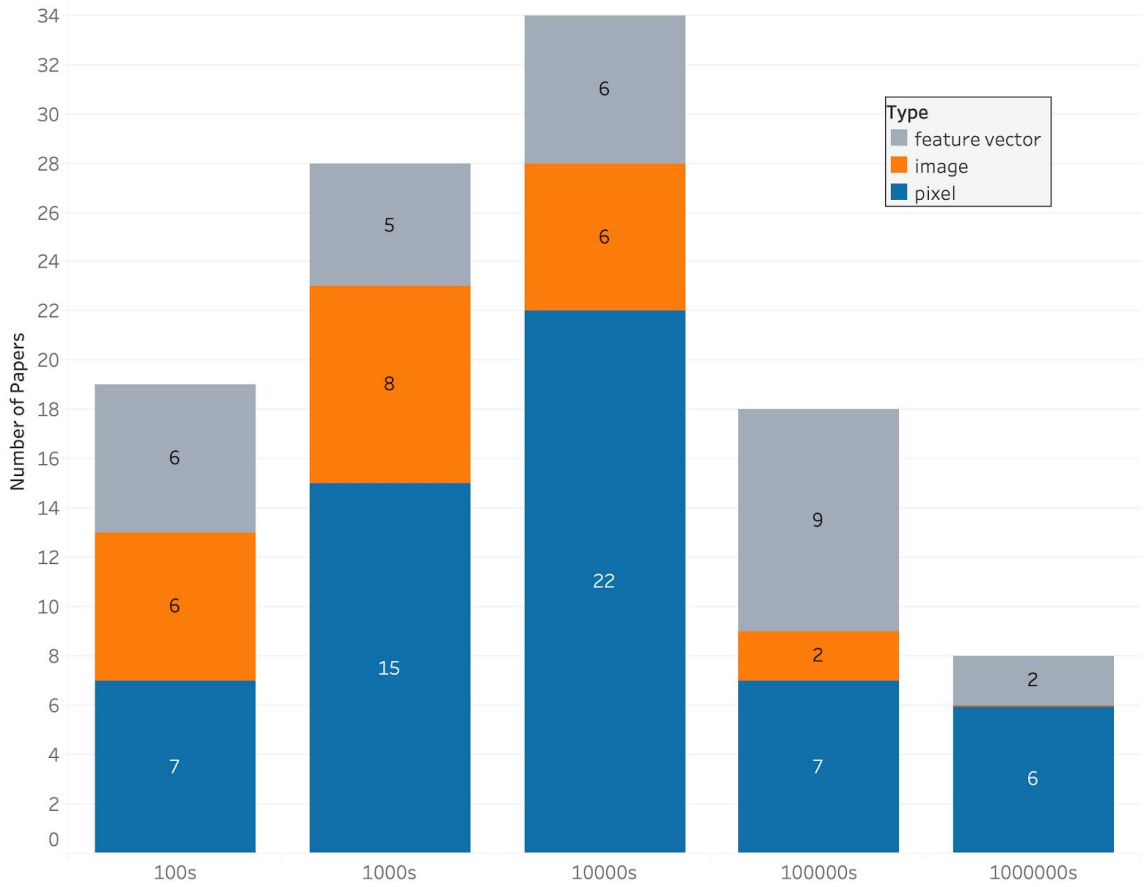


Application Area

ML in Earth Science: Current Landscape



Data from Instrument



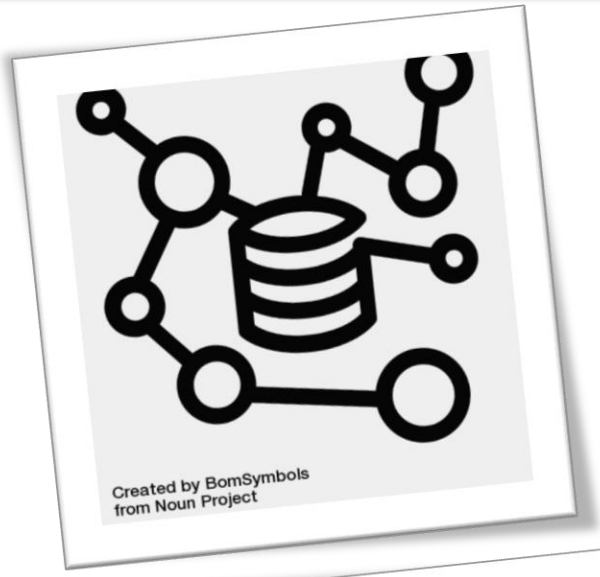
Training Data Size

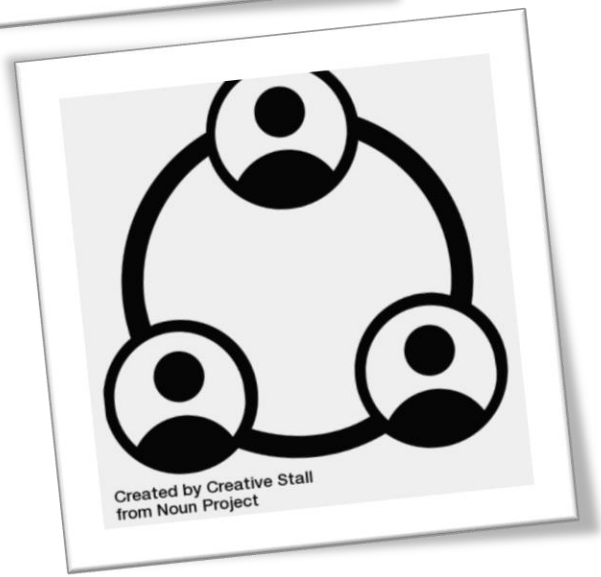
Why ML for Earth Science Data Systems?

- Exploit the large archives of Earth Science data
  - Open, large and wide variety of data from multiple instruments
- Collocated data and cloud platform enables scaling
- Augment and improve data systems operations (automation) and services (search/discovery)
- Opportunities for novel research and applications
  - ML models show promise of outperforming established benchmarks in prediction, forecasting, classification, and recommendations.

Challenges in applying ML for Earth Science

- Data Issues
  - Training dataset
    - Scarcity, Scale, Quality
    - SMEs dependency
  - Data Complexity
    - Use of data from authoritative sources
    - Integration of heterogeneous data sources (resolution, projection etc.)
- Collaborations (Team Science)
  - ML expertise to select the correct ML algorithm/architecture
  - Domain expertise for problem-specific optimization
  - ML model interpretation





ML Road Map: Six Focus Areas

Address Training Data Scarcity

- Leverage current archive

Community building

- Strategic partnerships

Practical ML Applications

- Operations and applications

Catalog of ML Models

New ML Tools

- Leverage existing data services

Operational Infrastructure

- Infusion of models into operations