The PDS Discipline Namespace for Machine Learning Products

Mystery

Kiri L. Wagstaff[1], Michael McAuley[1], and Minh Le[2]

[1]Jet Propulsion Laboratory, California Institute of Technology; [2]Stripe

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

The PDS Discipline ~~Namespace~~ Sub-model for Machine Learning Products

Mystery

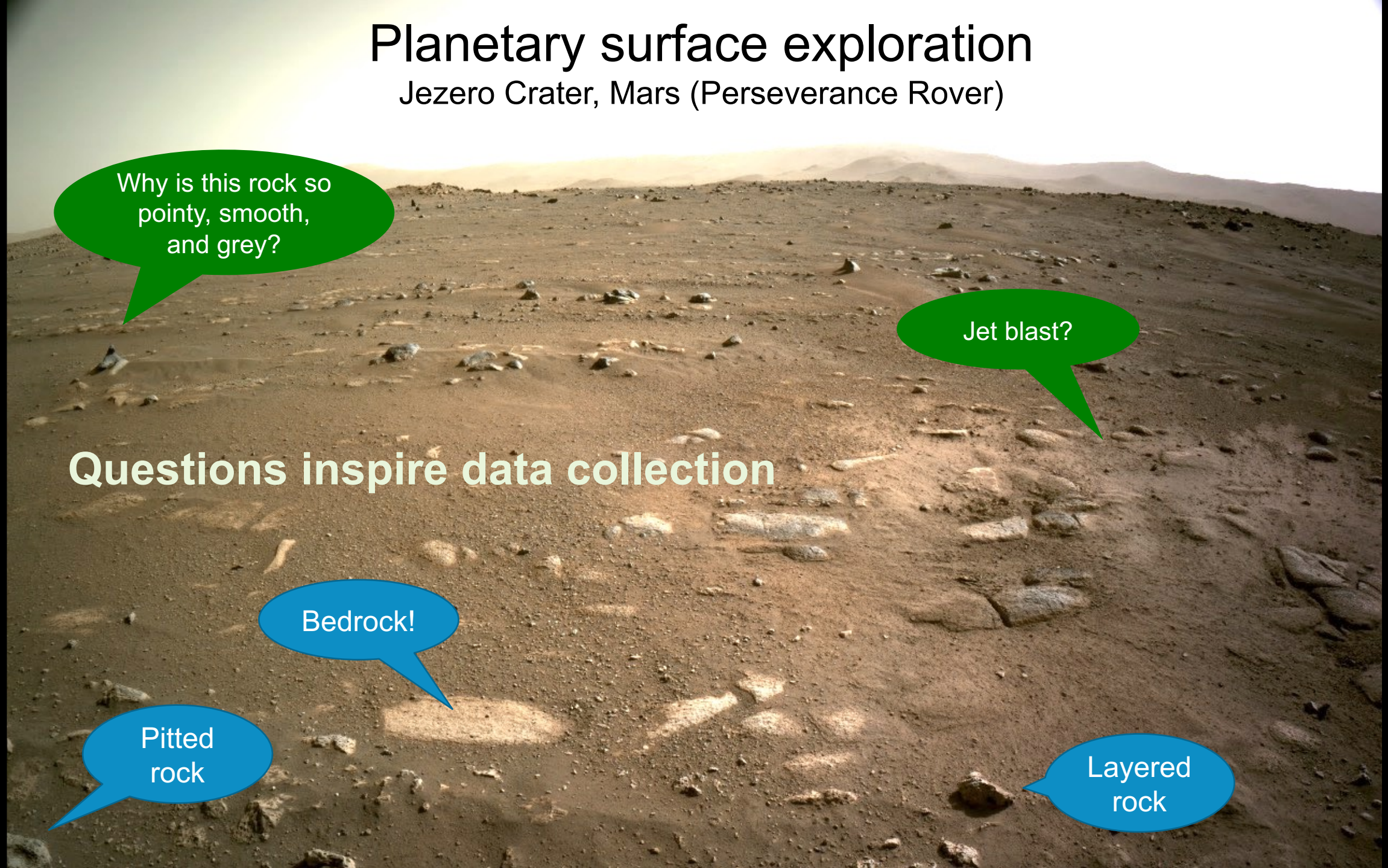
Kiri L. Wagstaff[1], Michael McAuley[1], and Minh Le[2]

[1]Jet Propulsion Laboratory, California Institute of Technology; [2]Stripe

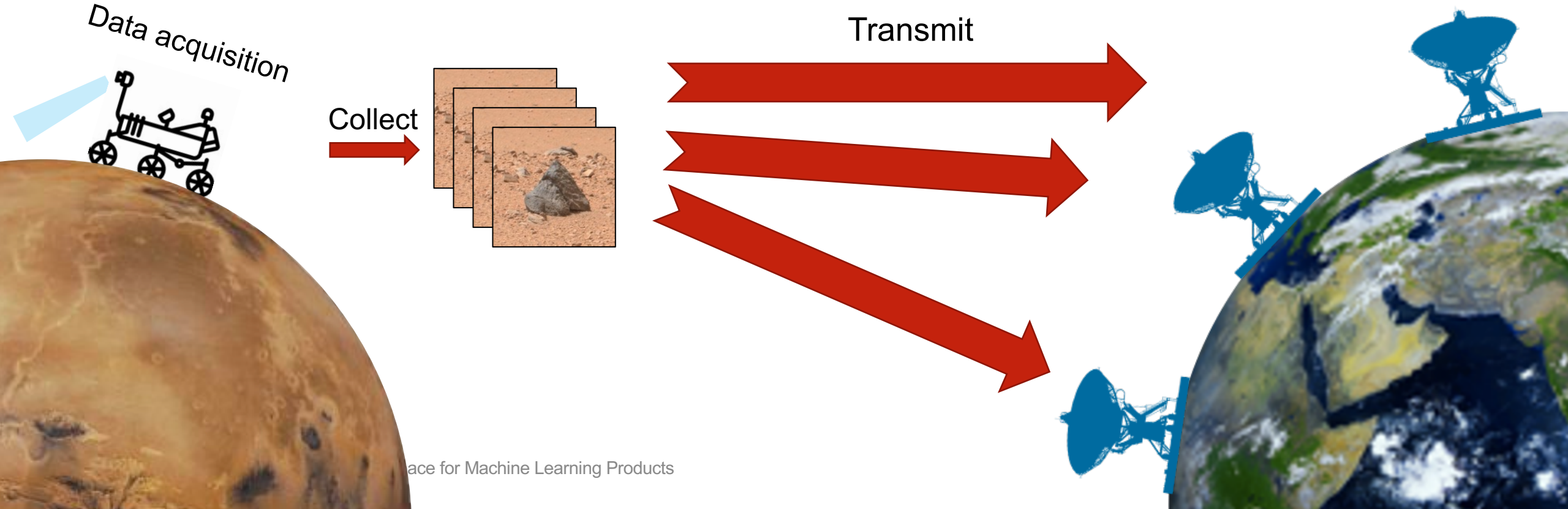
This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.

Planetary surface exploration

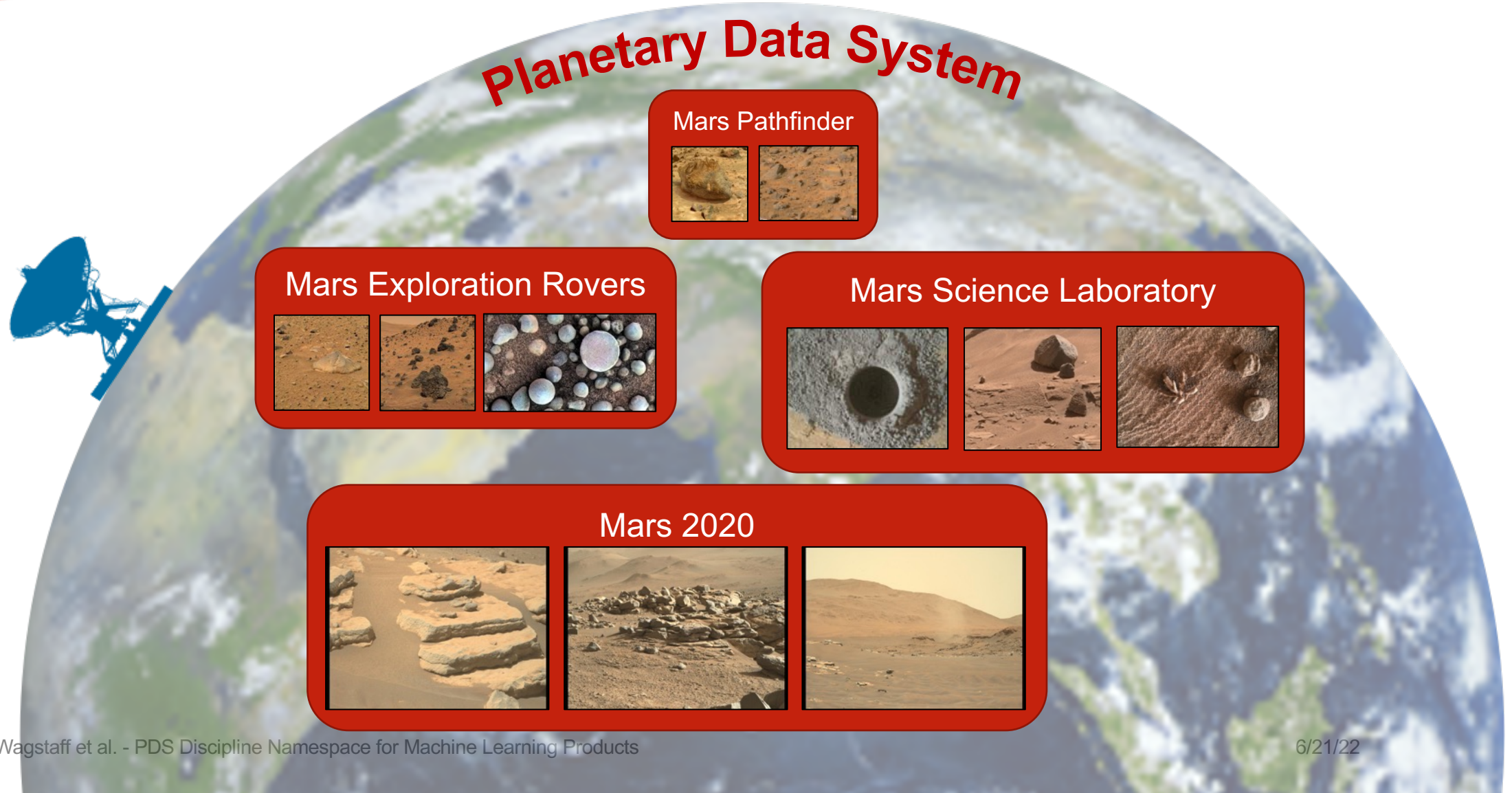
Jezero Crater, Mars (Perseverance Rover)



Planetary data collection…



Planetary data collection and archiving



Planetary Data System (PDS) metadata

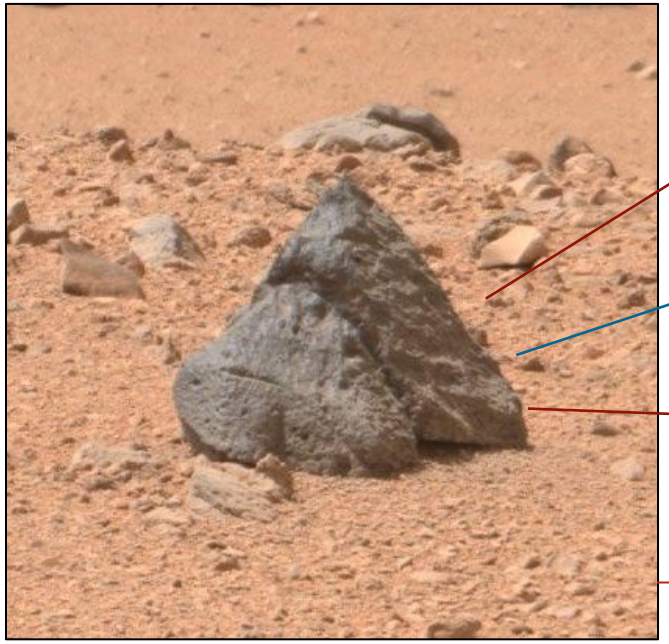
- Who collected it?
- What is it?
- When was it collected?
- Where was it collected?
- Why was it collected?

PDS4 Data Label (XML)

- Identification Area
  - Logical identifier
  - Version id
  - …
- Observation Area
  - Observing system
  - Target id
  - Time
  - Location
  - Rationale
  - …

Terms defined by PDS4 Information Model

What about derived products?



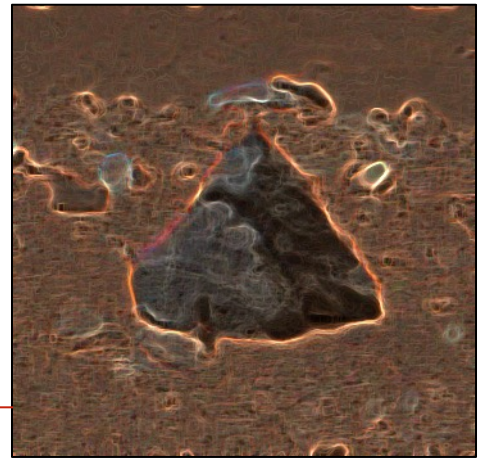
Classification

Float rock!

Basalt!

Sandstone!

Novelty map



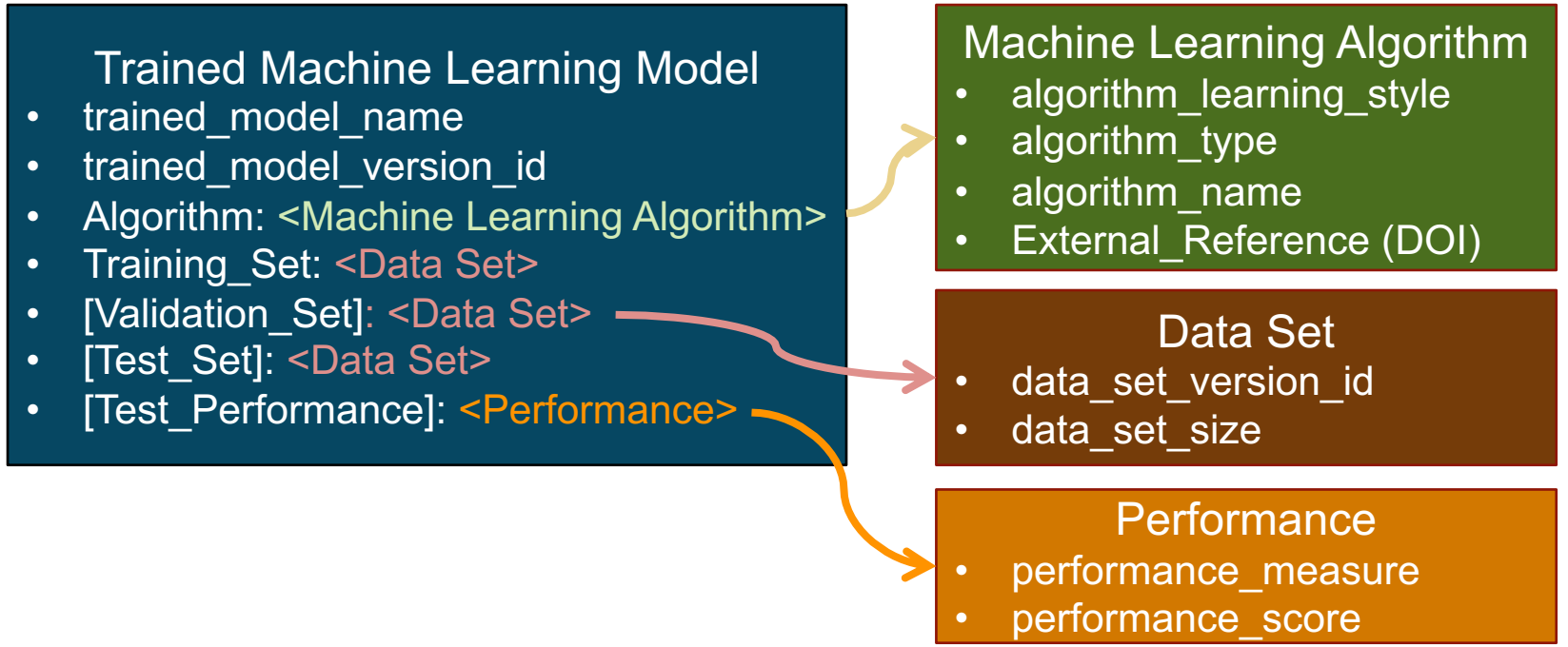
- Machine learning analysis of products generates new information
- These derived products generate new questions
  - Who created the derived product?
  - How was it done? (What model or process was used?)
  - Etc.

Documenting a model

- Model cards for model reporting [Mitchell et al., 2019]
  - Creator, date, version, type, training algorithm(s), paper
  - How to cite it, license for use, contact for questions/comments, intended use
  - Training and evaluation data
  - Metrics and performance measures
  - Ethical considerations, caveats, and recommendations
- Benefits
  - Reproducibility
  - Ease of use by others
  - Provenance: tracing information to its source

Machine Learning Analysis Discipline Sub-model

- Extends the PDS4 Information Model
- Metadata structure and keywords to define the ML model that generated the product



More information: https://github.com/pds-data-dictionaries/ldd-ml

Example: MSLNet Classifier Product

MSL Image Archive

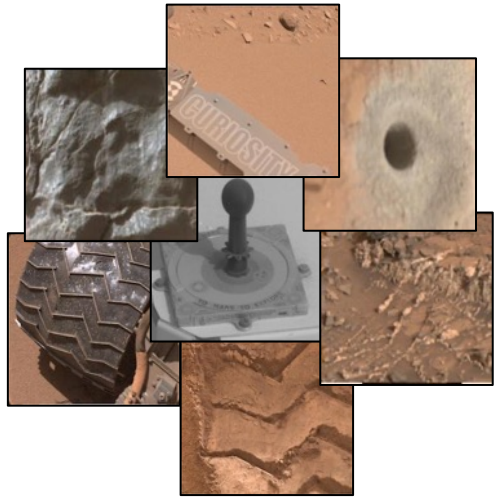


Image Classifier (15 classes)

ML product (predictions for 1M images)

# image,	class,	probability
image1,	close-up rock,	0.95
image2,	float rock,	0.92
image3,	wheel tracks,	0.97
…		

Describe

PDS4 Data Label (XML)

- Identification Area
  - Logical identifier
  - Version id
  - …
- Observation Area
  - …
- **Machine Learning Area**
  - Machine Learning Algorithm (CNN, supervised)
  - Image Data Set (version 2, 5920 train, 300 validation, 600 test)
    - Source: Mastcam and MAHLI
  - …

MSL Image Archive

Benefits of the ML Discipline Sub-model

- ML products can be archived as regular products in the PDS, increasing visibility of ML products
- Discipline sub-model provides structure for future ML product deliveries (e.g., PDART)
- New ML metadata terms enable traceability/provenance to increase understanding and use of ML products
- PDS users can search/filter ML products by ML model attributes of interest

- We invite your ideas and requests for extensions!
  - New Steward for the Machine Learning Discipline Submodel: Sara Bond (sara.a.bond@jpl.nasa.gov)

PDS Planetary Data System

Thank you: PDS Cartography and Imaging Sciences Node