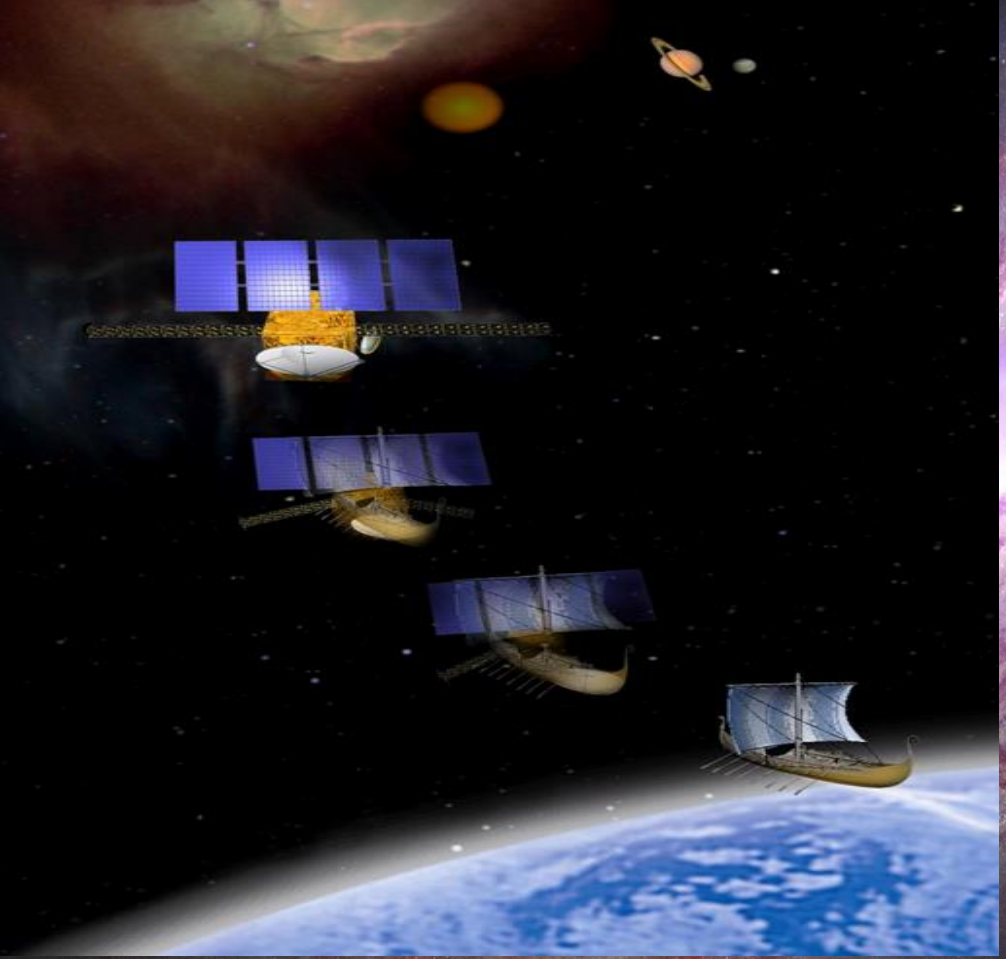


Data Science and Computing on the Path to Autonomy

Jet Propulsion Laboratory
California Institute of Technology

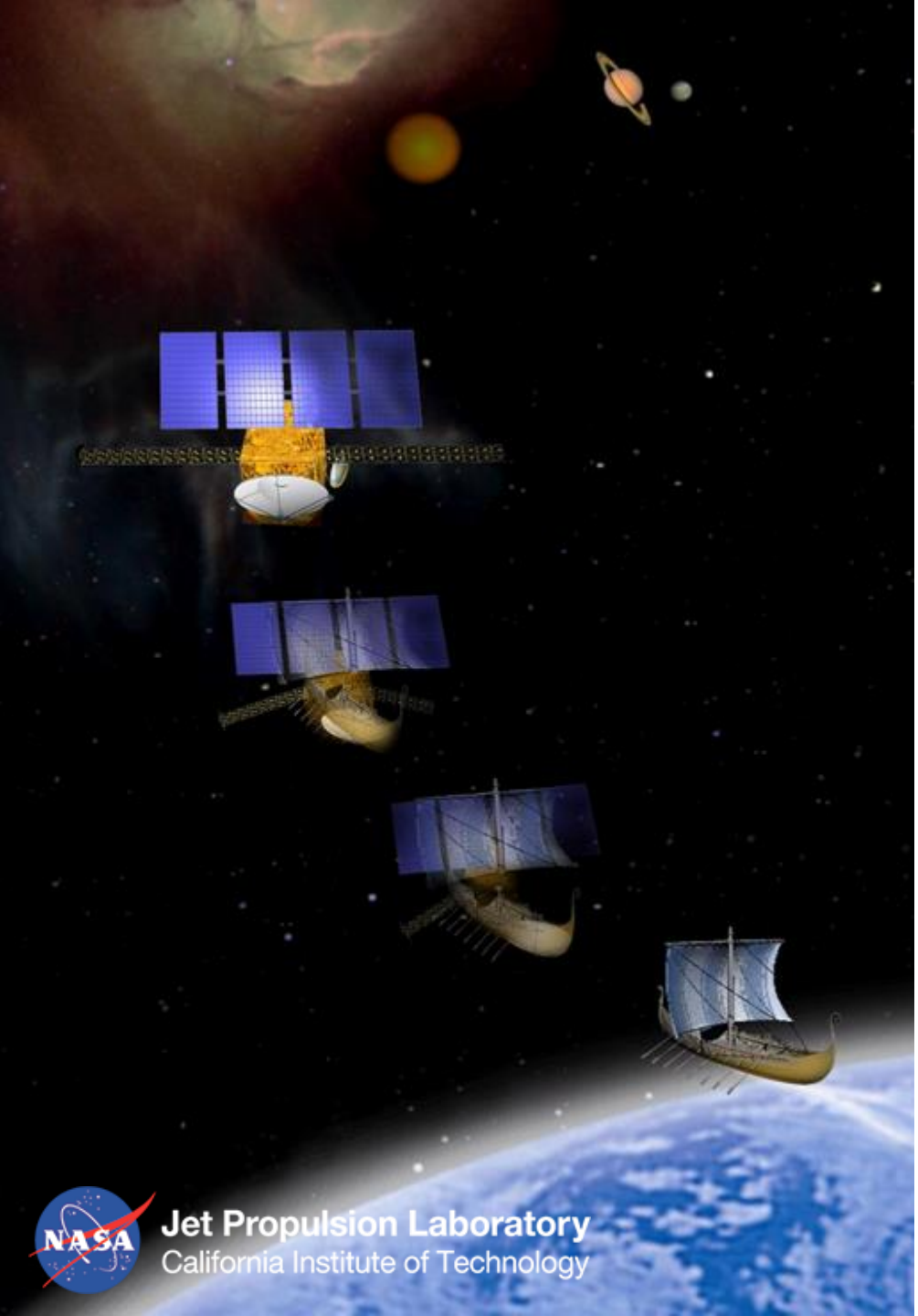
Dr. Richard J. Doyle

SMC-IT 2017

September 29, 2017

Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement by the United States Government or the Jet Propulsion Laboratory, California Institute of Technology.

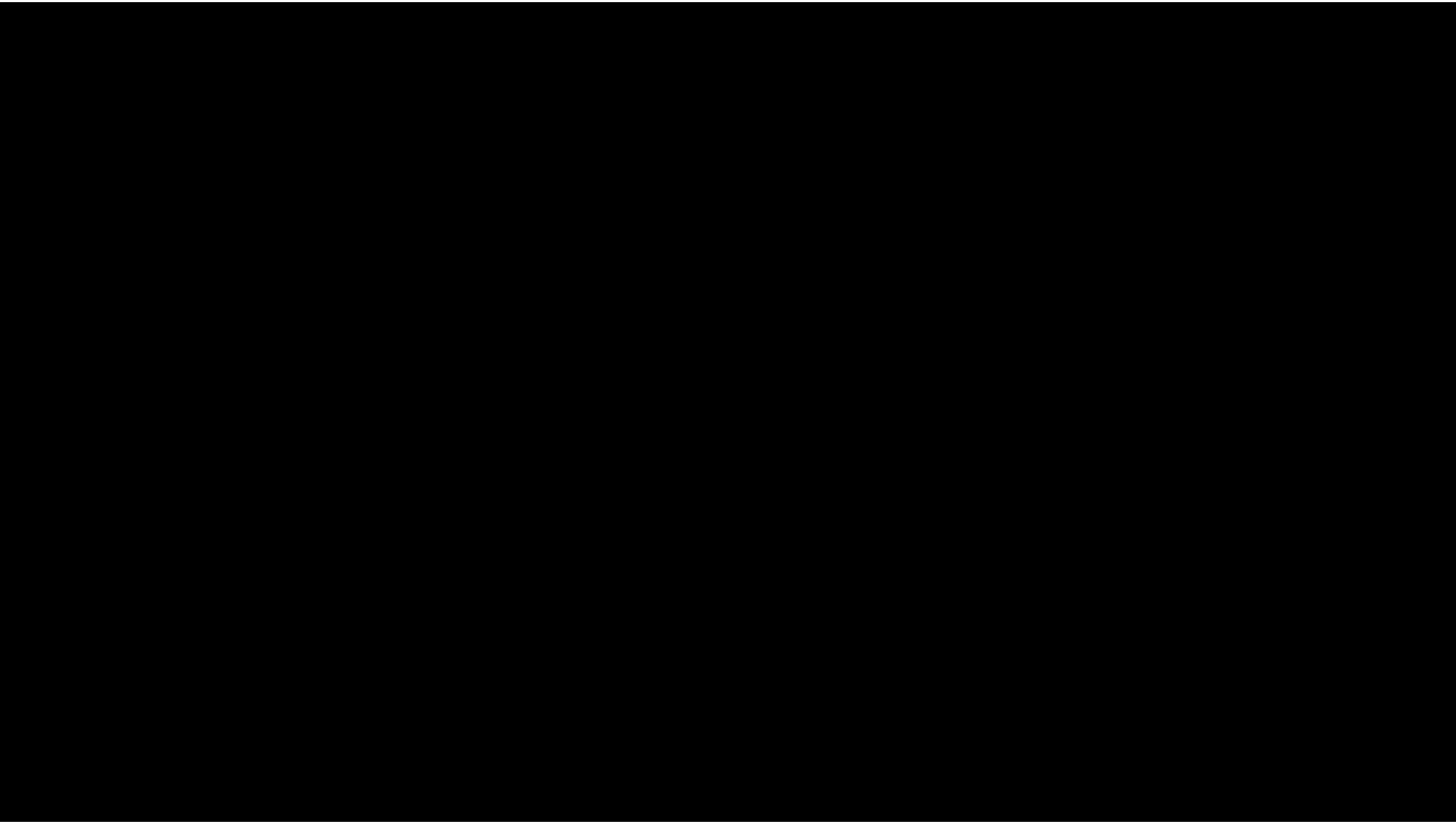
© 2017 California Institute of Technology. Government sponsorship acknowledged.



JPL

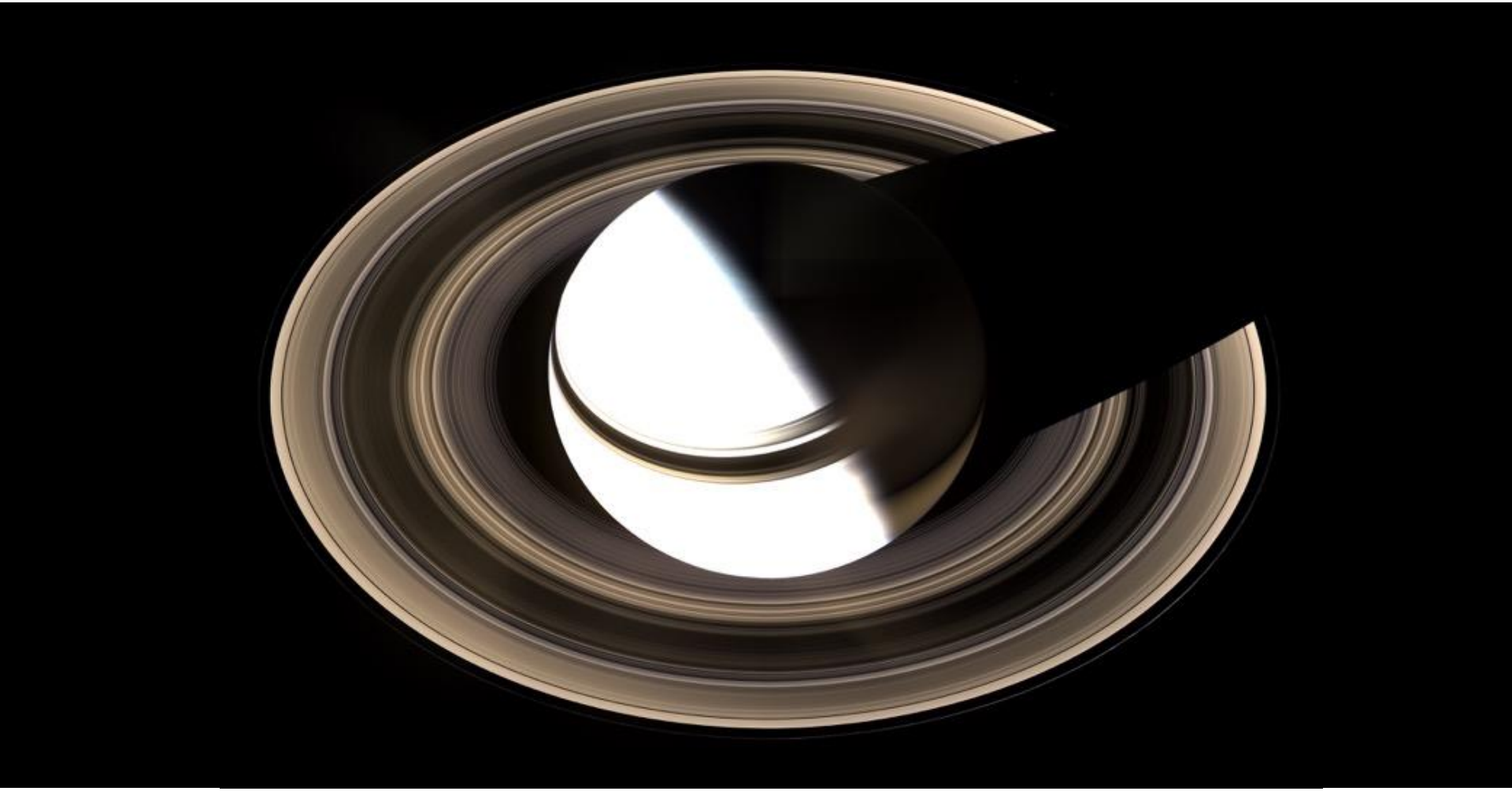
Dare Mighty Things

A Culture of Innovation for Decades

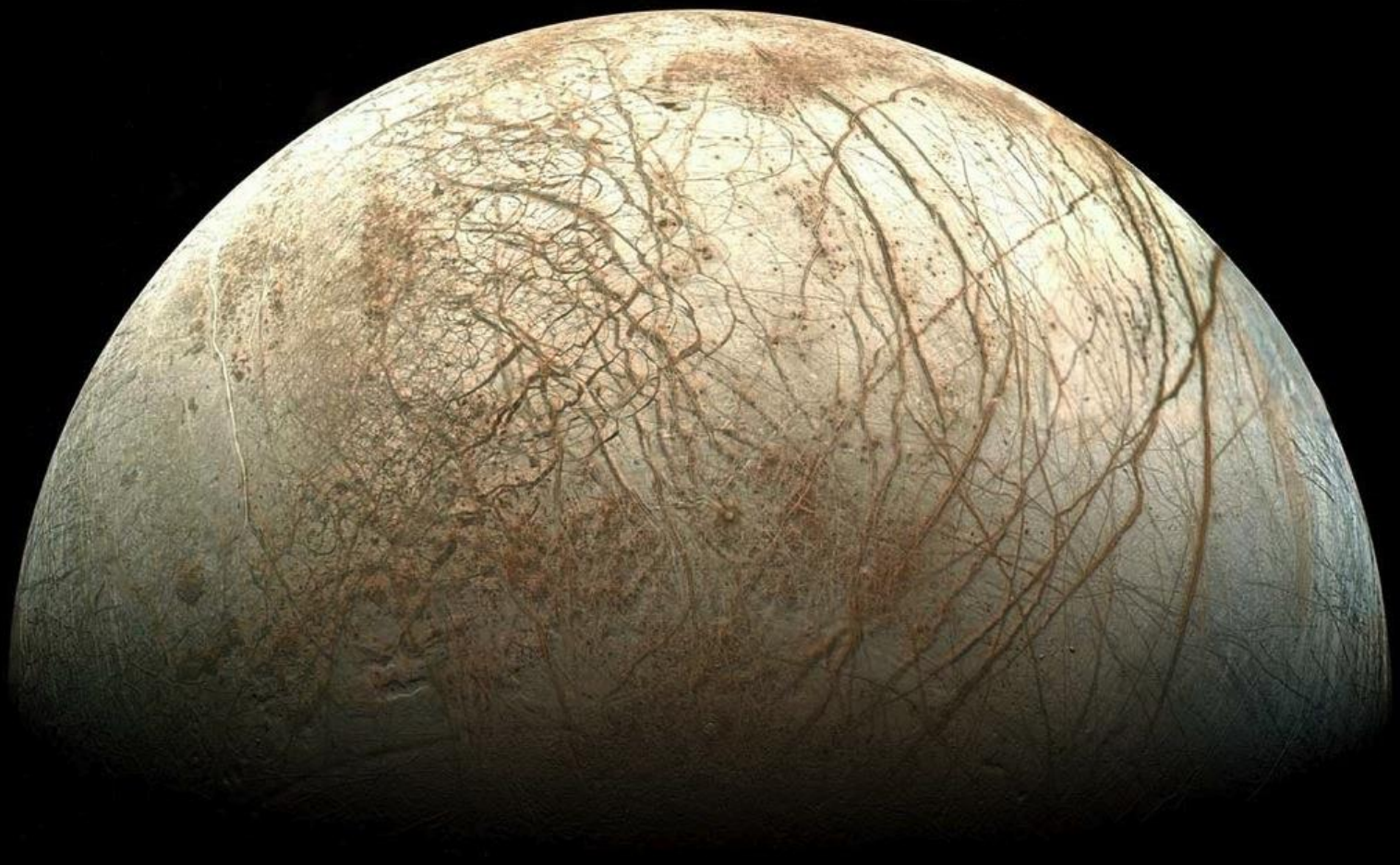


Cassini

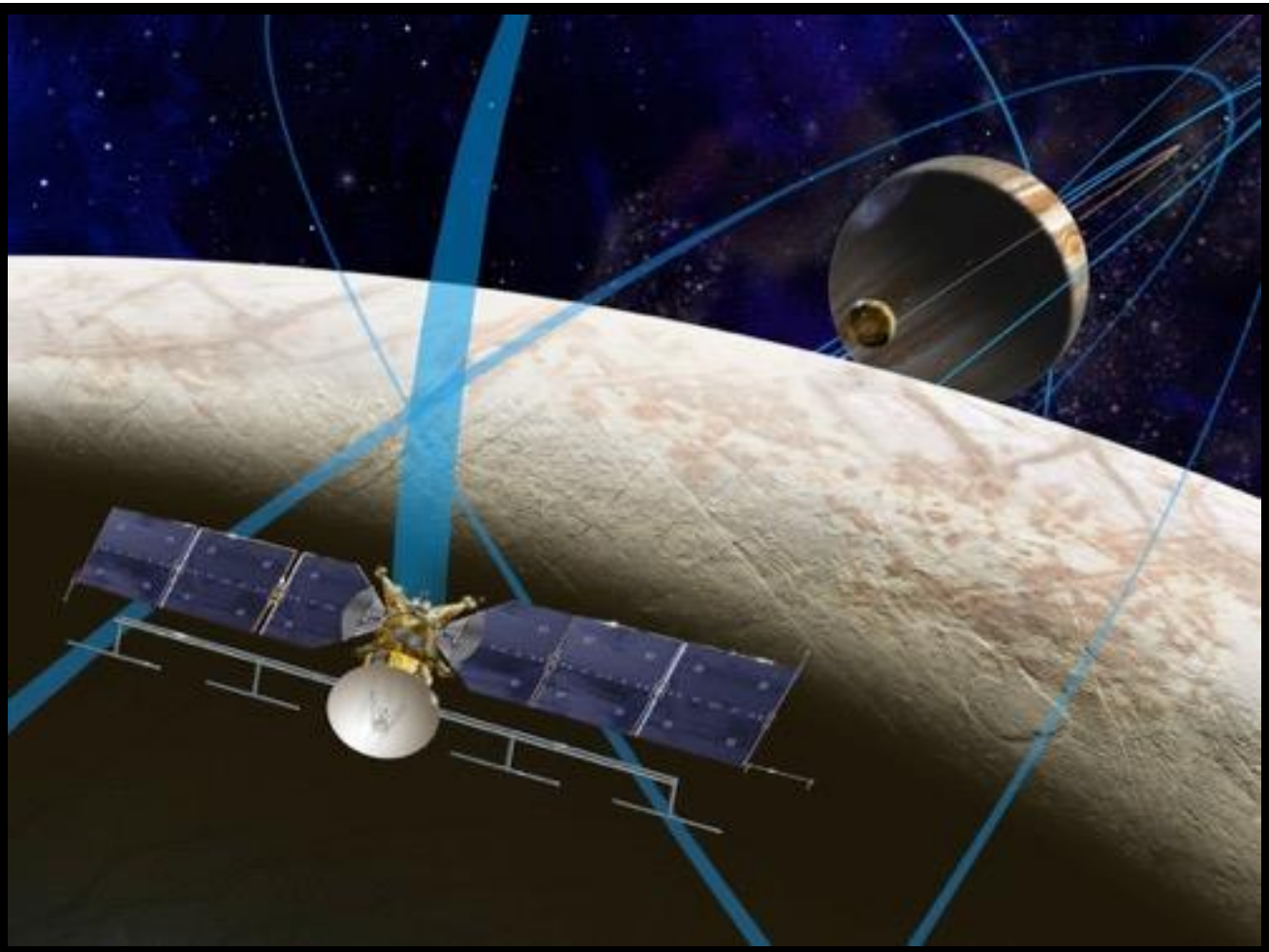
Orbit Insertion June 28, 2004
Grand Finale September 15, 2017



Europa: Gem of the Jupiter System

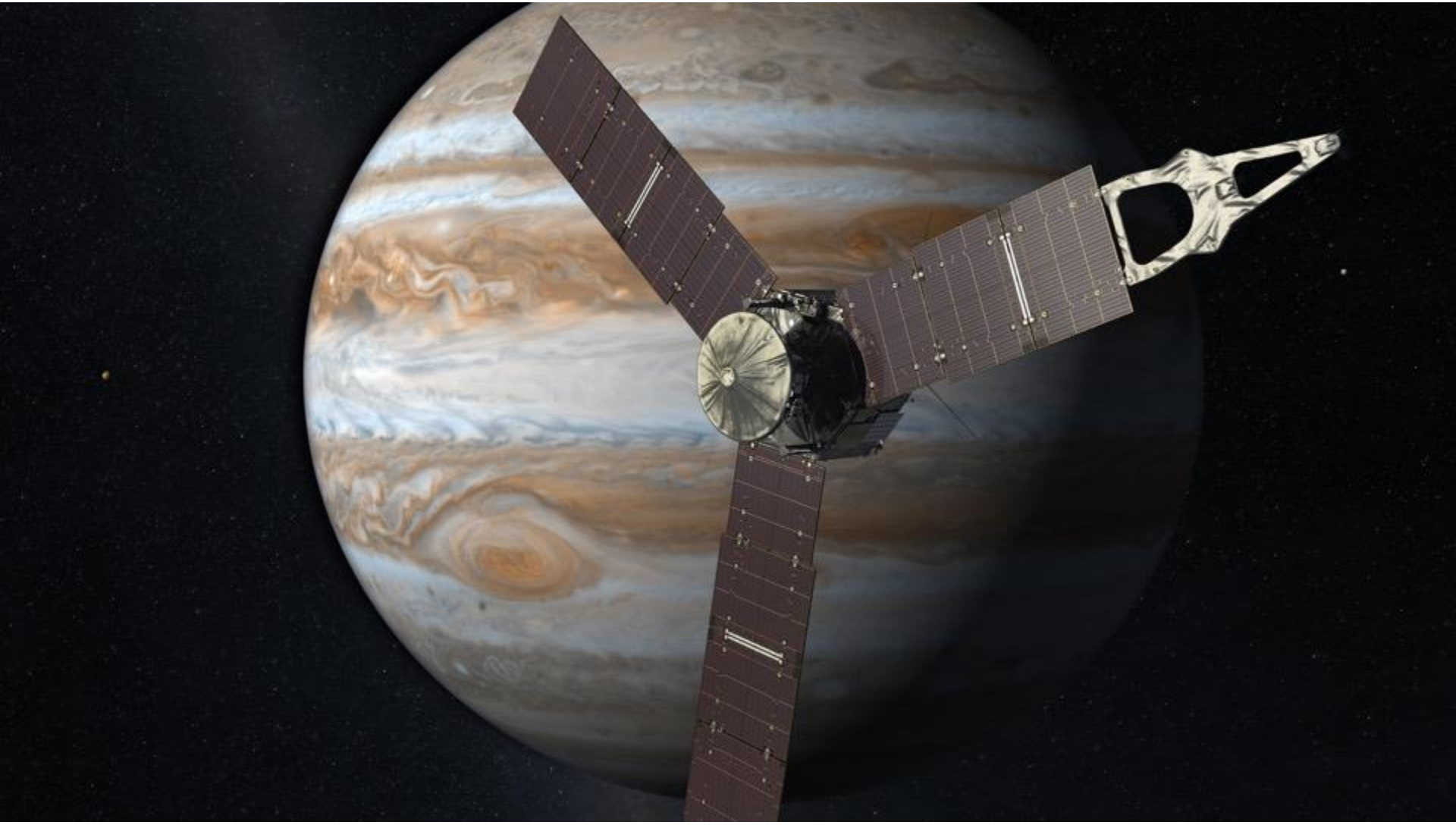


Europa Clipper



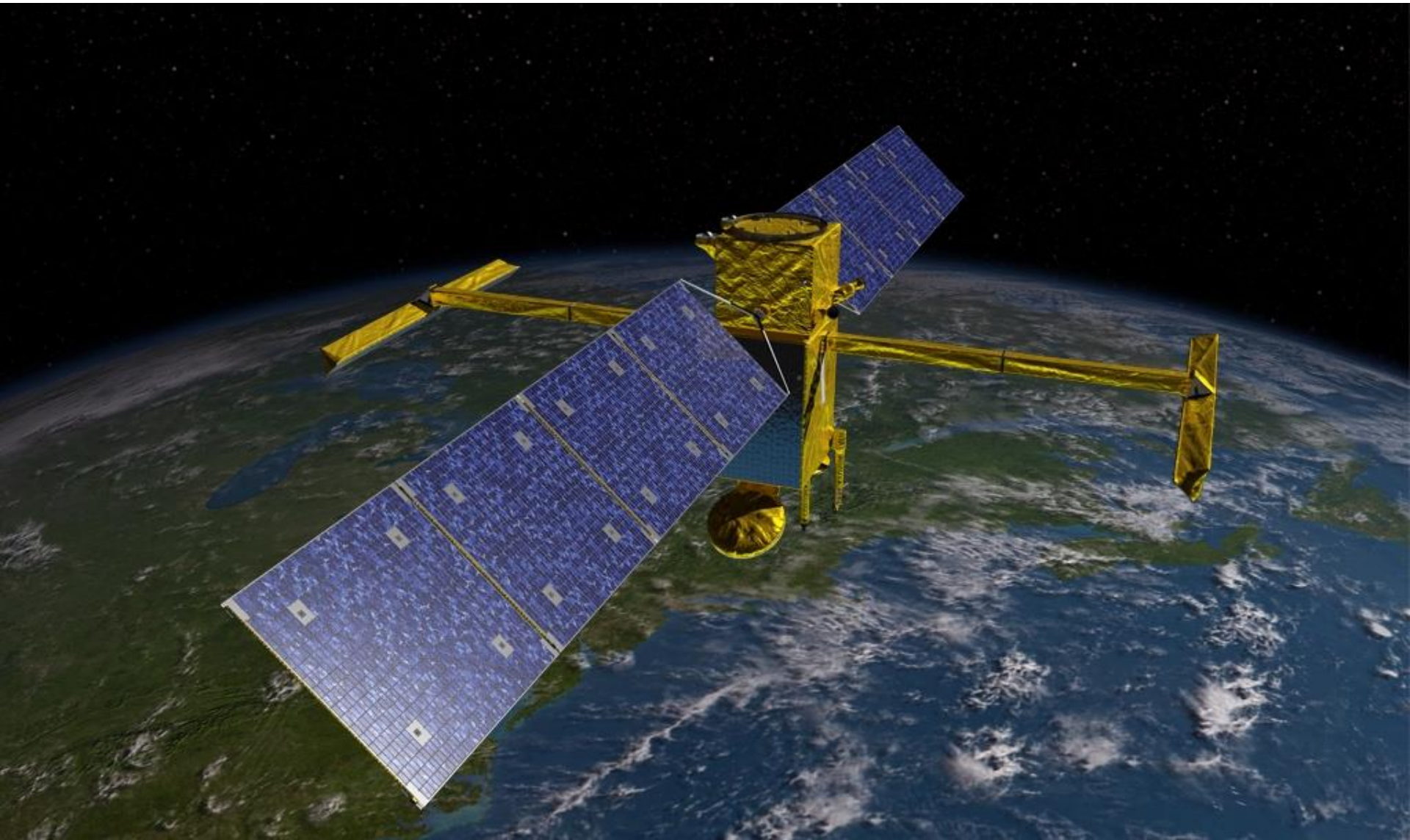
Juno

Orbit Insertion July 4, 2016



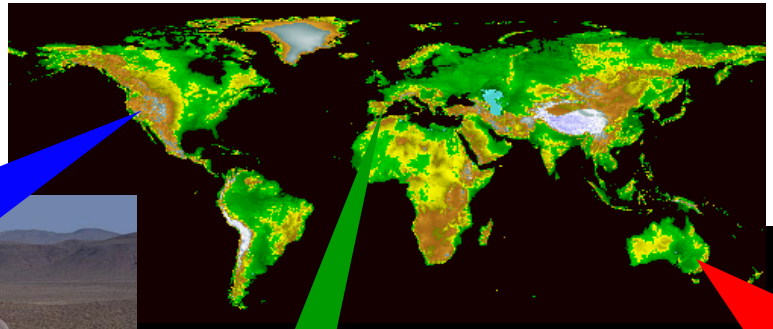
Surface Water Ocean Topography

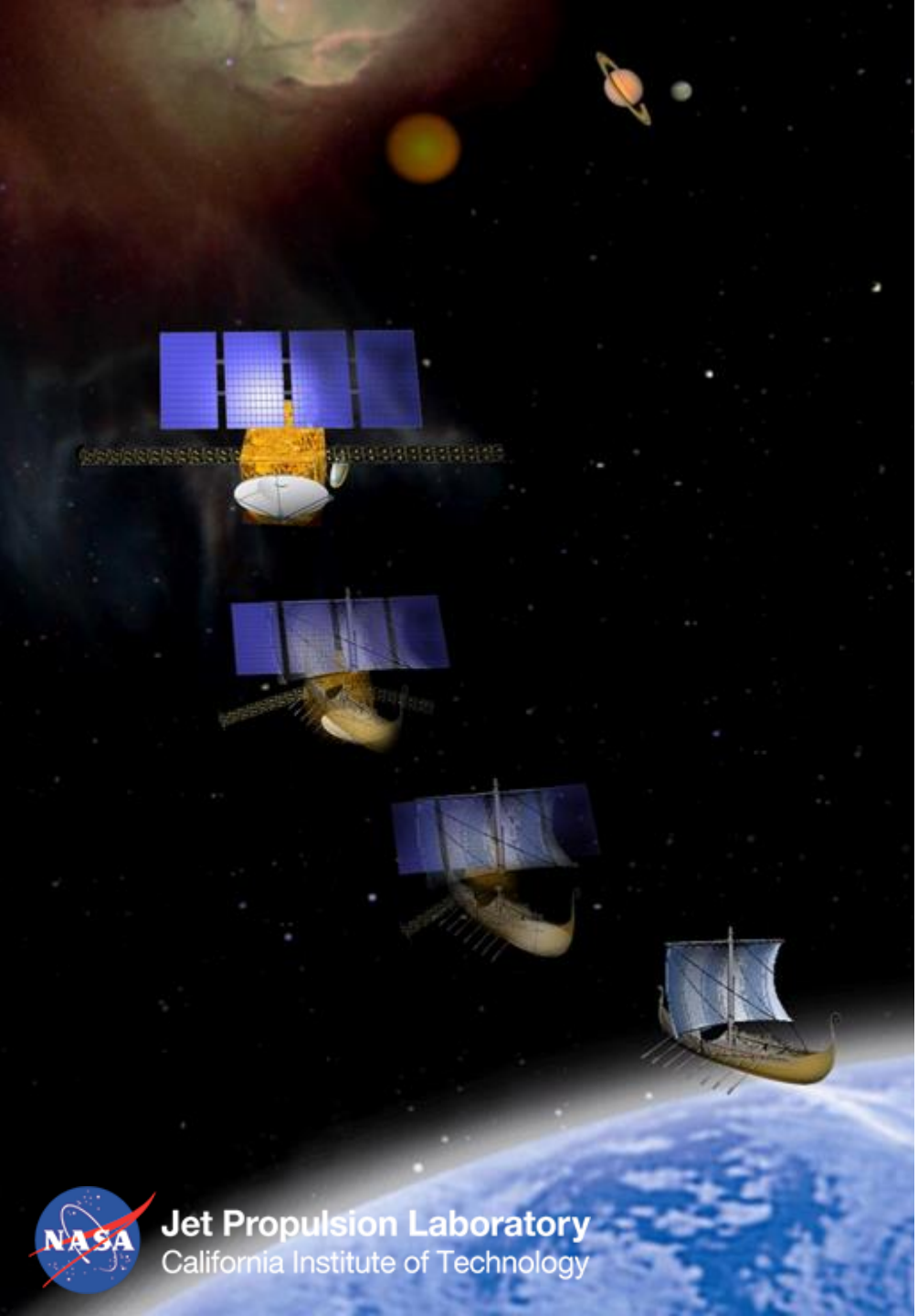
2020 Launch



Jet Propulsion Laboratory
California Institute of Technology

Deep space exploration enabled by NASA's Deep Space Network (DSN)





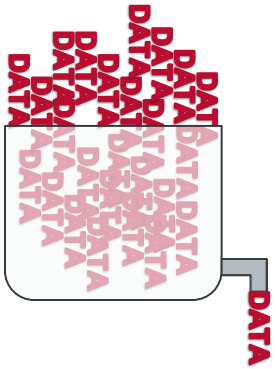
The Challenge

Enabling Data-Driven Exploration

Terms: Big Data and Data Science

Big Data

When needs for data collection, processing, management and analysis go beyond the capacity and capability of available methods and software systems



Data Science

S*calable* architectural approaches, techniques, software and algorithms which alter the paradigm by which data is collected, managed and analyzed

Data Lifecycle Model

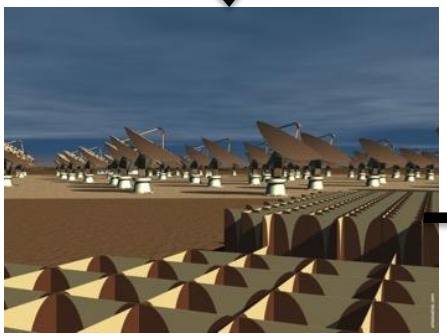
for NASA Space Missions

Emerging Solutions
- *Onboard Data Data Analytics*
- *Onboard Data Prioritization*
- *Flight Computing*

(1) Too much data, too fast; cannot transport data efficiently enough to store

Observational Platforms and Flight Computing

Emerging Solutions
- *Intelligent Ground Stations*
- *Agile MOS-GDS*



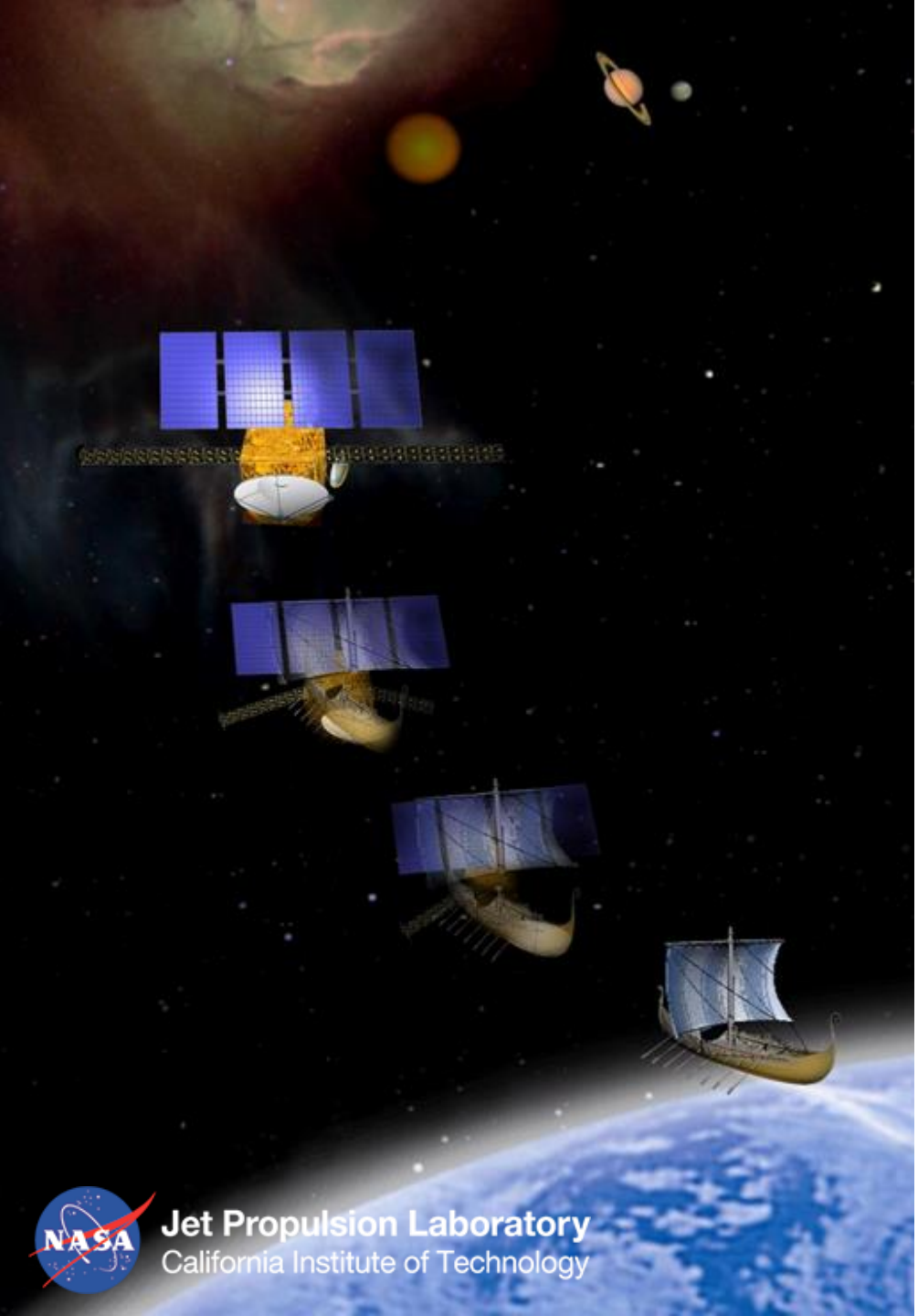
(2) Data collection capacity at the instrument continually outstrips data transport (downlink) capacity

Ground-based Mission Systems

Massive Data Archives and Big Data Analytics

Emerging Solutions
- *Data Discovery from Archives*
- *Distributed Data Analytics*
- *Advanced Data Science Methods*
- *Scalable Computation and Storage*

(3) Data distributed in massive archives; many different types of measurements and observations



Data-Driven Flight Systems

The Future of Flight Computing

Increasing Computing Capability Onboard

Heading Toward Multicore in Space

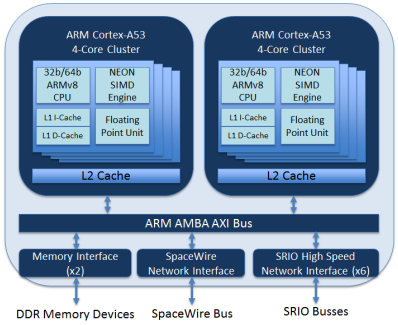
Voyager computer

- 8,000 instructions/sec and kilobytes of memory

iPhone

- 14 billion instructions/sec and gigabytes of memory

Curiosity (Mars Science Laboratory)
Processor: 200 MOPS BAE RAD750



HPSC (NASA STMD / AFRL)
Processor: 15 GOPS, extensible

HPSC Use Cases

NASA Mission Applications

Human Spaceflight (HEOMD) Use Cases

1. Cloud Services
2. Advanced Vehicle Health Management
3. Crew Knowledge Augmentation Systems
4. Improved Displays and Controls
5. Augmented Reality for Recognition and Cataloging
6. Tele-Presence
7. Autonomous & Tele-Robotic Construction
8. Automated Guidance, Navigation, and Control (GNC)
9. Human Movement Assist

Science Mission (SMD) Use Cases

1. Extreme Terrain Landing
2. Proximity Operations / Formation Flying
3. Fast Traverse
4. New Surface Mobility Methods
5. Imaging Spectrometers
6. Radar
7. Low Latency Products for Disaster Response
8. Space Weather
9. Science Event Detection and Response
10. Immersive Environments for Science Ops / Outreach

High value and mission critical applications identified by NASA scientists and engineers

Credit: R. Doyle, W. Powell, R. Some

NASA Flight Computing Drivers

as Derived from Use Cases

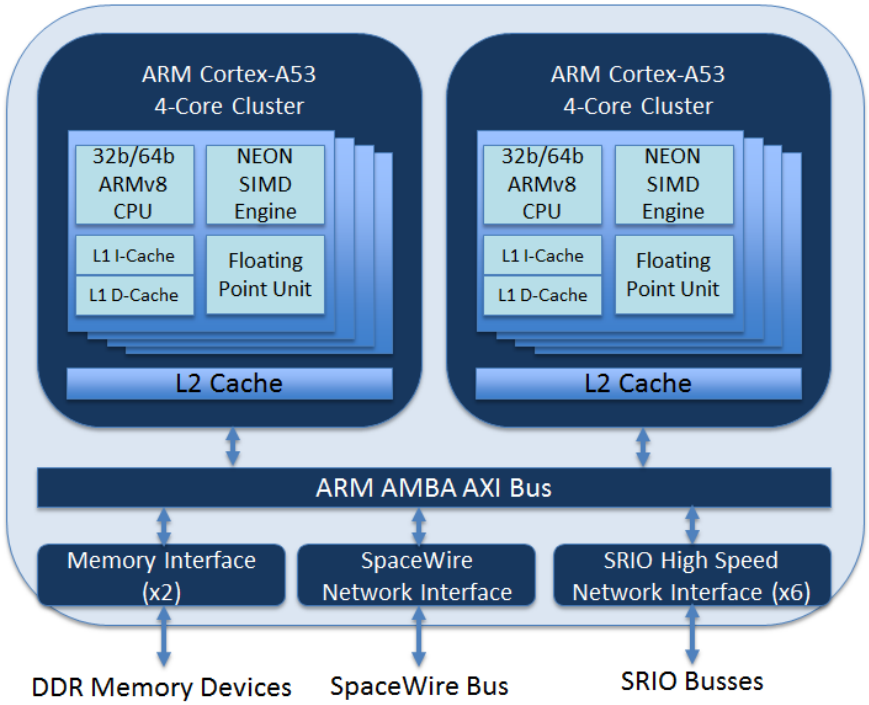
Computation Category	Mission Need	Objective of Computation	Flight Architecture Attribute	Processor Type and Requirements
Vision-based Algorithms with Real-Time Requirements	• Terrain Relative Navigation (TRN) • Hazard Avoidance • Entry, Descent & Landing (EDL) • Pinpoint Landing	• Conduct safe proximity operations around primitive bodies • Land safely and accurately • Achieve robust results within available timeframe as input to control decisions	• Severe fault tolerance and real-time requirements • Fail-operational • High peak power needs	**• Hard real time / mission critical** **• Continuous digital signal processing (DSP) + sequential control processing (fault protection)** **• High I/O rate** **• Irregular memory use** **• General-purpose (GP) processor (10's – 100's GFLOPS) + high I/O rate, augmented by co-processor(s)**
Model-Based Reasoning Techniques for Autonomy	• Mission planning, scheduling & resource management • Fault management in uncertain environments	• Contingency planning to mitigate execution failures • Detect, diagnose and recover from faults	• High computational complexity • Graceful degradation • Memory usage (data movement) impacts energy management	**• Soft real time / critical** **• Heuristic search, data base operations, Bayesian inference** **• Extreme intensive & irregular memory use (multi-GB/s)** **• > 1GOPS GP processor arrays with low latency interconnect**
High Rate Instrument Data Processing	High resolution sensors, e.g., SAR, Hyper-spectral	• Downlink images and products rather than raw data • Opportunistic science	• Distributed, dedicated processors at sensors • Less stringent fault tolerance	**• Soft real time** **• DSP/Vector processing with 10-100's GOPS (high data flow)** **• GP array (10-100's GFLOPS) required for feature ID / triage**

Future NASA use cases require dramatic improvement over the RAD750

Credit: R. Doyle, W. Powell, R. Some

HPSC Reference Architecture

8-Core Extensible Chiplet

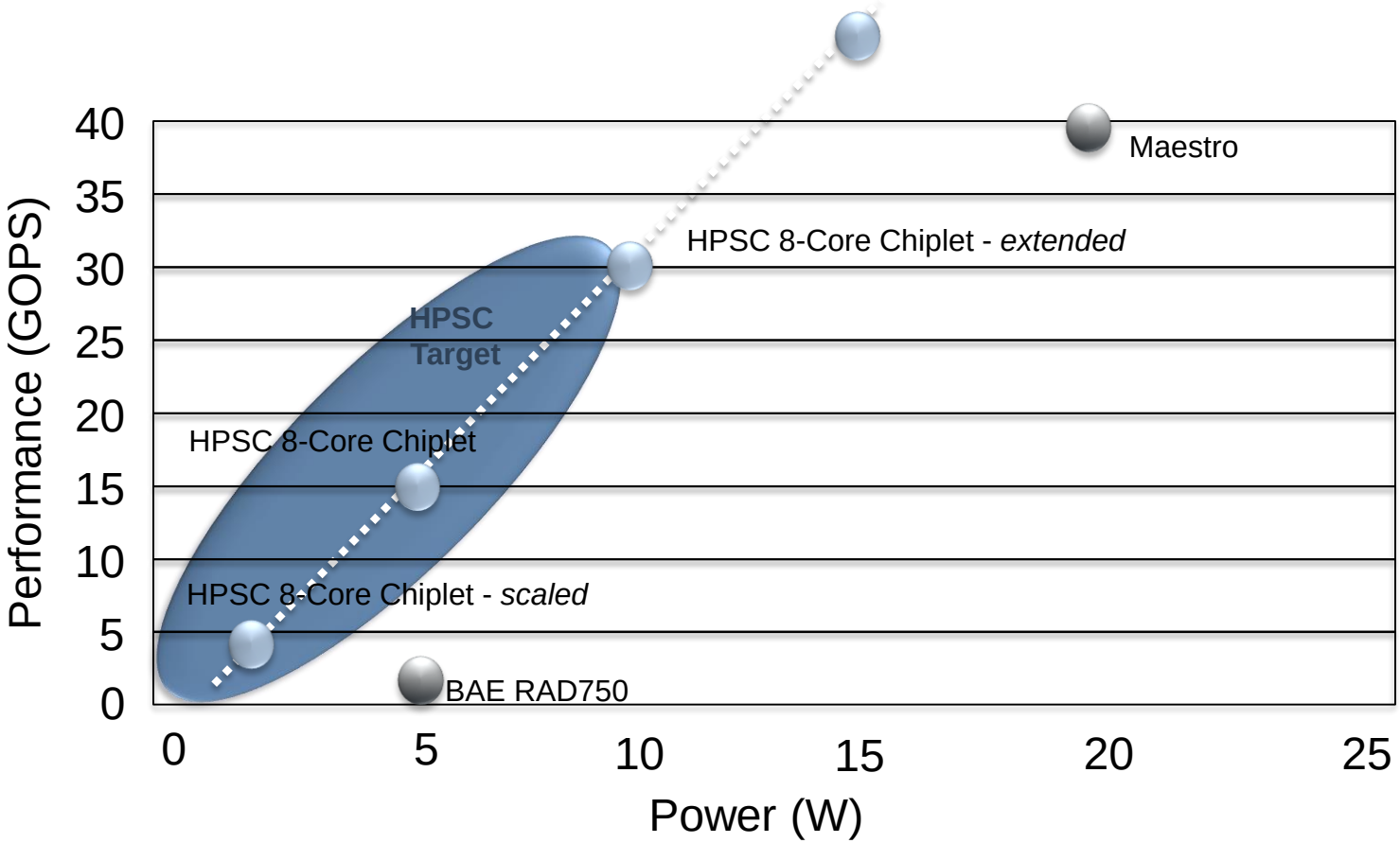


- A53 clusters for high bandwidth processing provide ~15 GOPS
- Typical device power is ~7 Watts (depending on memory and I/O bandwidth)
- Power Management – unused cores are dynamically de-powered or put to sleep
- Extensible, interoperable, flexible and future-proofed processor architecture enabling custom mission configurations

Credit: R. Doyle, W. Powell, R. Some

HPSC Game Changing Advance

Performance vs. Power



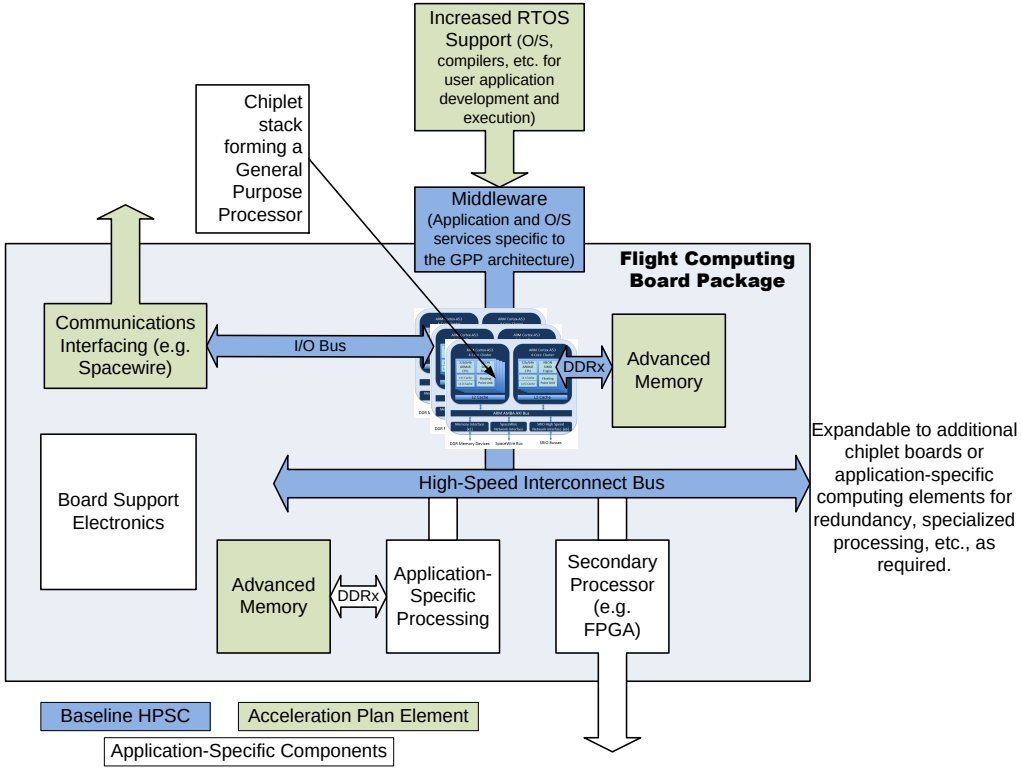
Extensible and interoperable, with trades enabled among performance, power and fault tolerance

Credit: R. Doyle, W. Powell, R. Some

HPSC

Ecosystem Concept

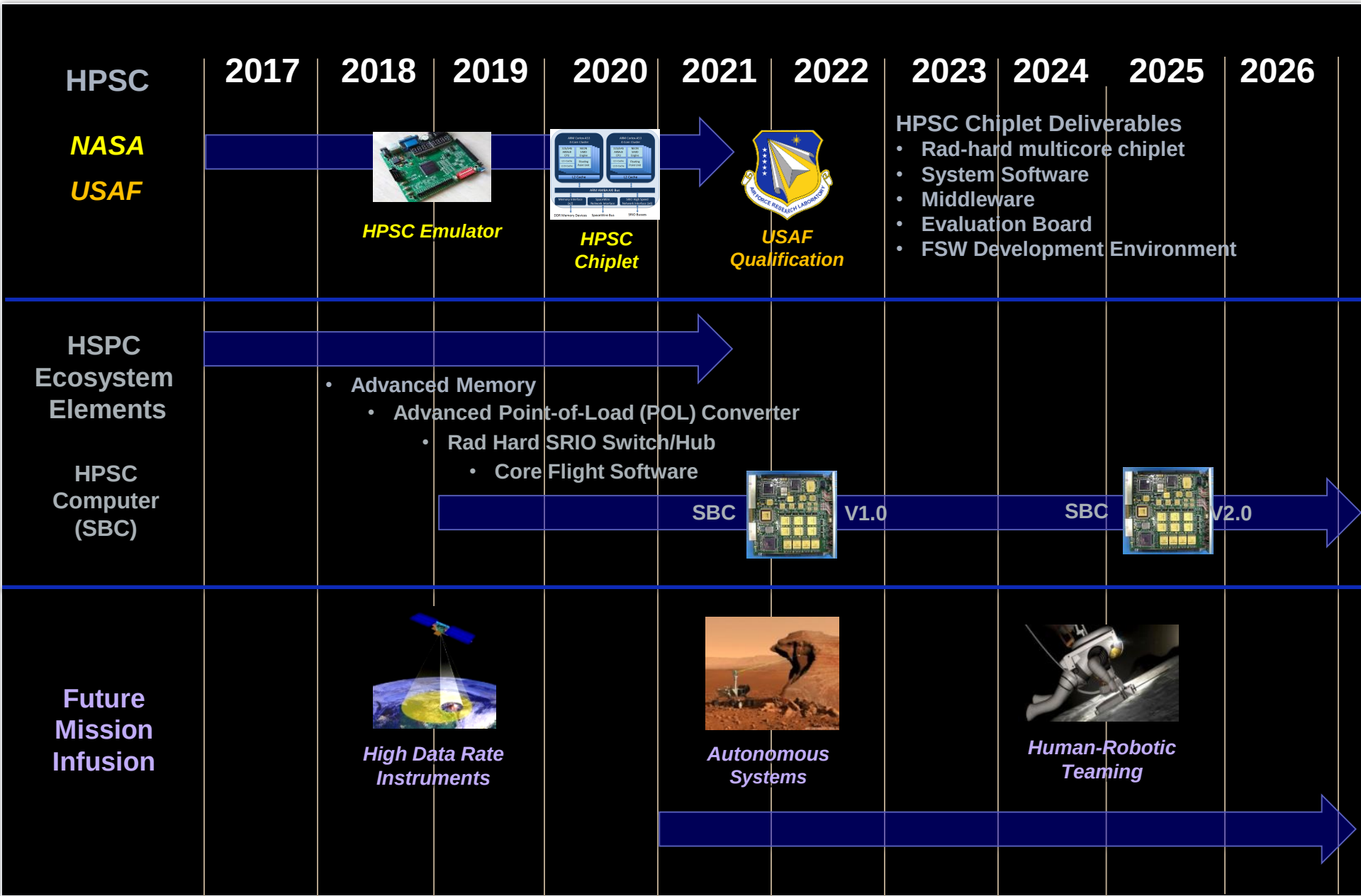
- Example of *HPSC* in a “full up” computing configuration where the *HPSC* “chiplet” is the General Purpose Processor that utilizes memory, support processing, and data communications interfaces

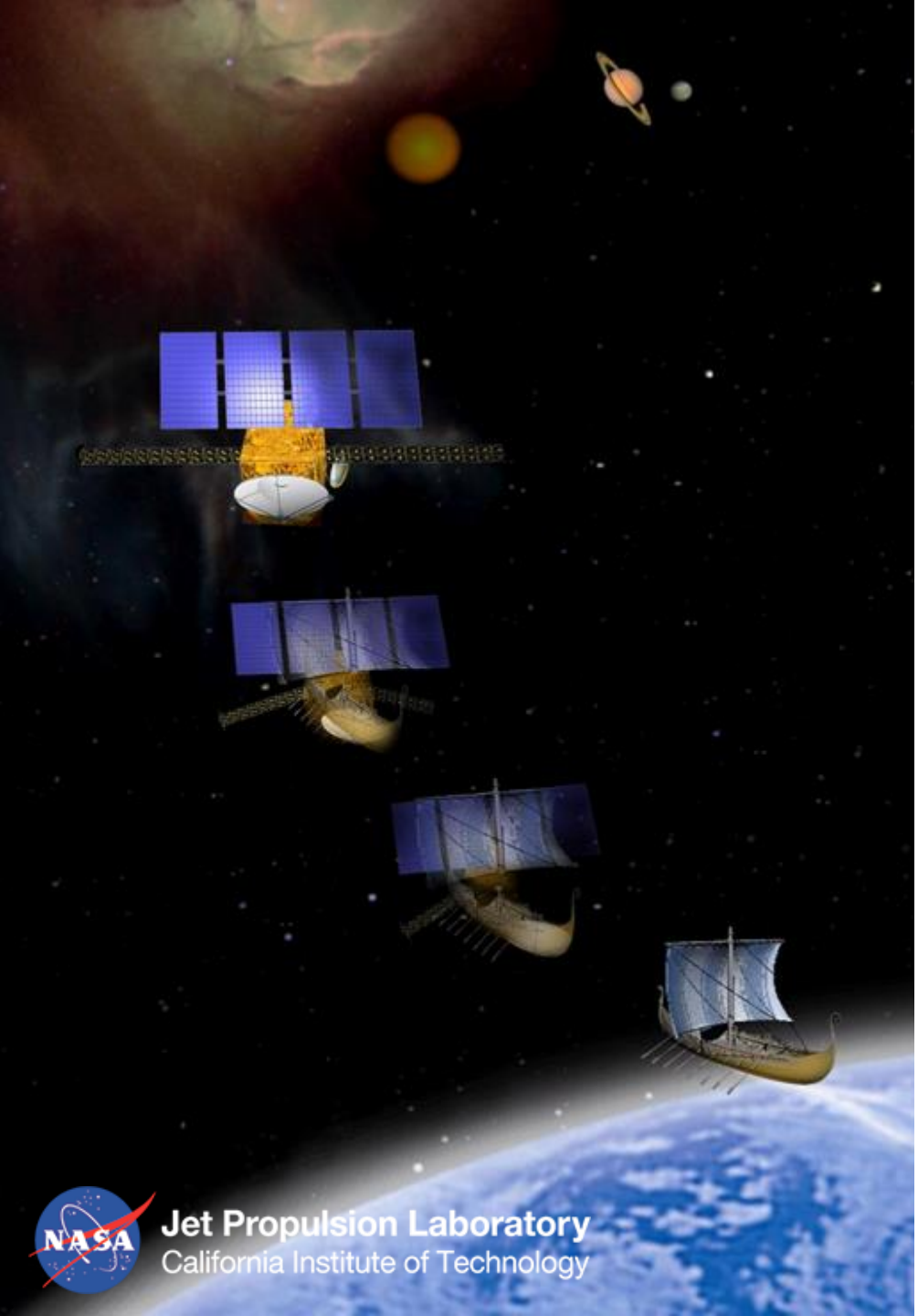


Credit: R. Some

HPSC

Alignment and Infusion Strategy





Data-Driven Flight Systems

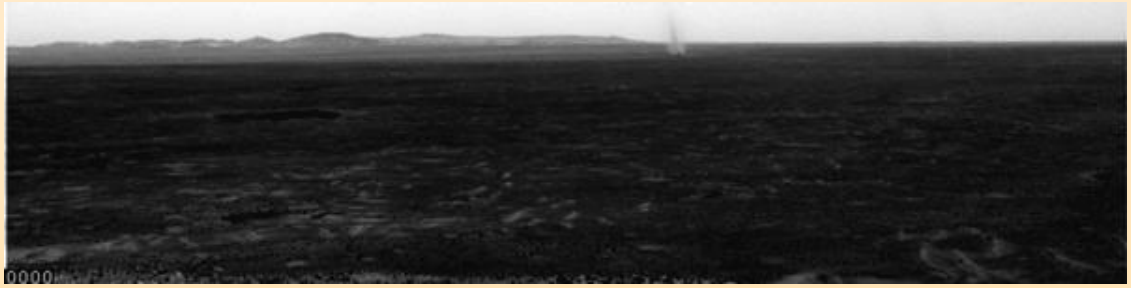
Moving Toward Onboard Analytics

Onboard Data Product Generation

Dust Devils on Mars

Dust devils are scientific phenomena of a transient nature that occur on Mars

- They occur year-round, with seasonally variable frequency
- They are challenging to reliably capture in images due to their dynamic nature
- Scientists accepted for decades that such phenomena could not be studied in real-time



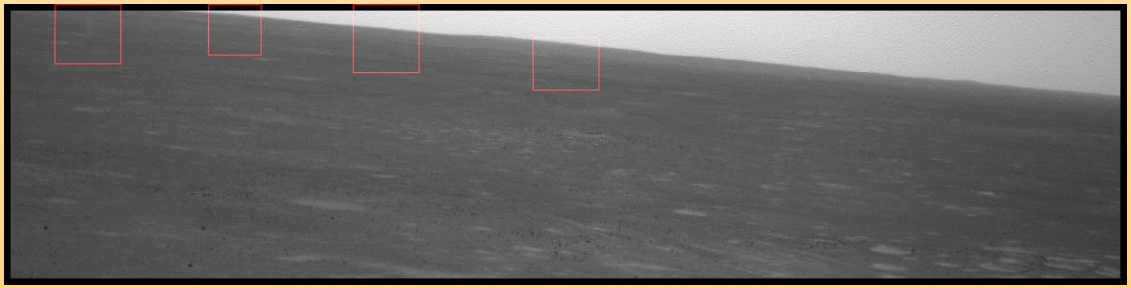
Spirit Sol 543
(July 13, 2005)

New onboard Mars rover capability (as of 2006)

- Collect images more frequently, analyze onboard to detect events, and only downlink images containing events of interest

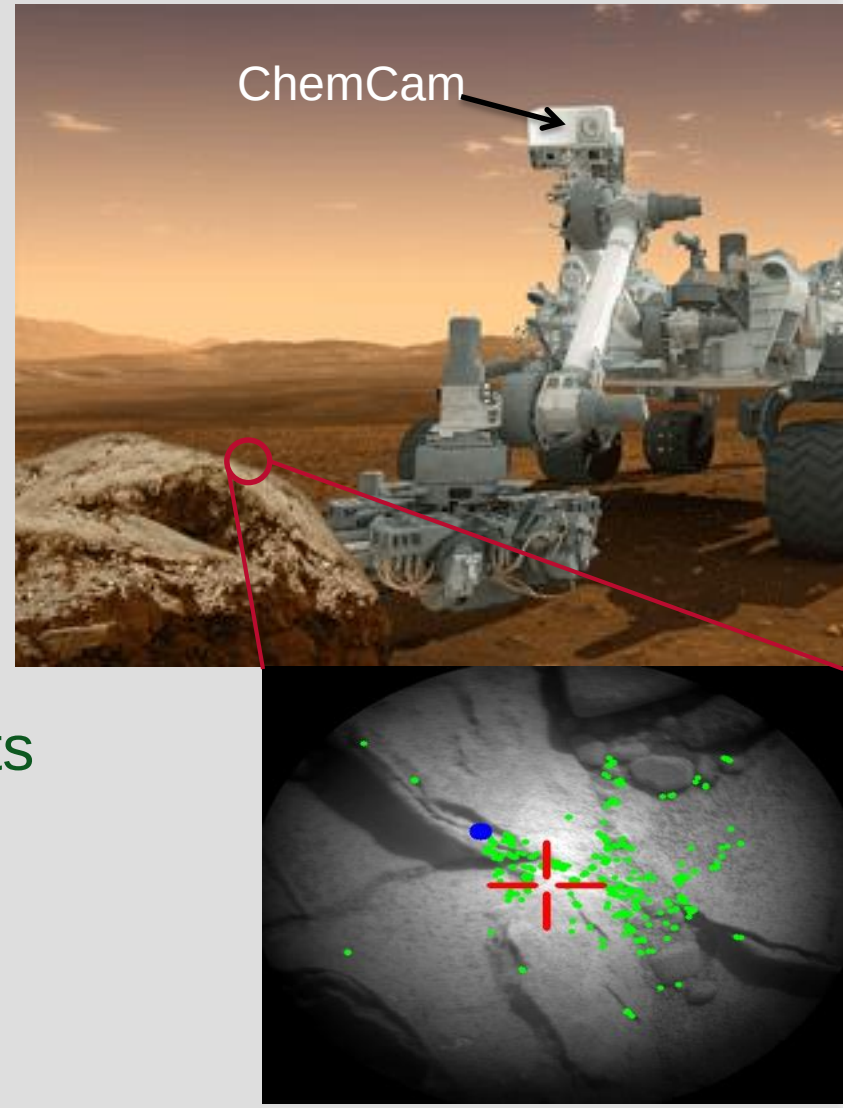
Benefit

- < 100% accuracy can dramatically increase science event data returned to Earth
- *First notification includes a complete data product*



Credit: T. Estlin. B. Bornestein, A. Castano, J. Biesiadecki, L. Neakrase, P. Whelley, R. Greeley, M. Lemmon, R. Castano, S. Chien and MER project team

AEGIS/ChemCam on MSL: Autonomous Explorations on Mars



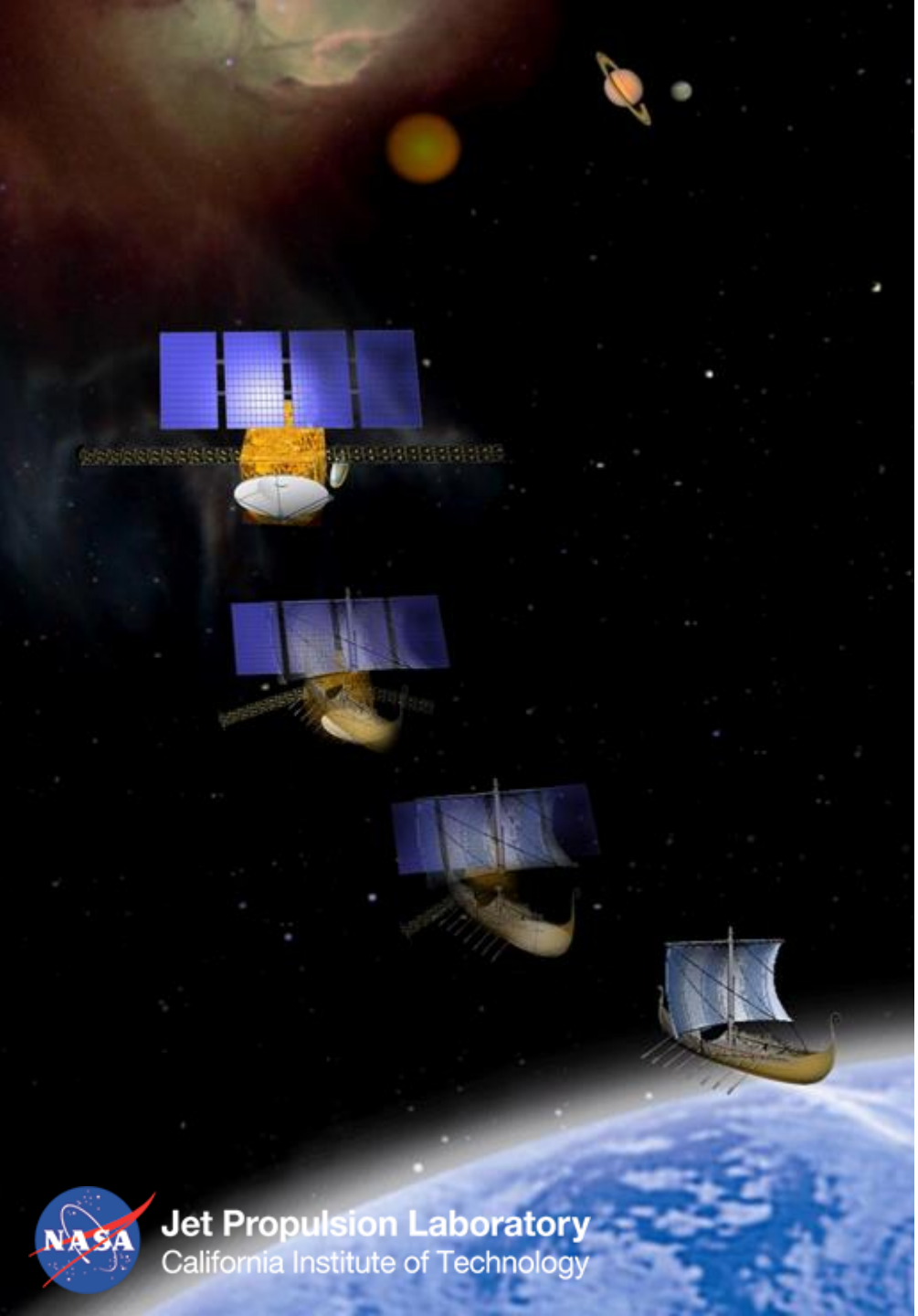
- Target & Zap Rock
- Manually Scheduled Targets

- Round Trip Delays
- Trouble hitting 1st time
- Targeted science not possible right after drive

- Autonomy selects interesting targets
- Refines targeting automatically

- ~30-100% additional ChemCam science targets on drive sols

Credit: T. Estlin, G. Doran, R. Francis et al



Data-Driven Ground Systems

Enabling Interactive Analytics

Opportunities for Data-Driven Ground Systems

Intelligent Ground Stations

Emerging Solutions
- *Anomaly Detection*
- *Combining DSN & Mission Data*
- *Attention Focusing*
- *Controlling False Positives*

Data-Driven Discovery from Archives

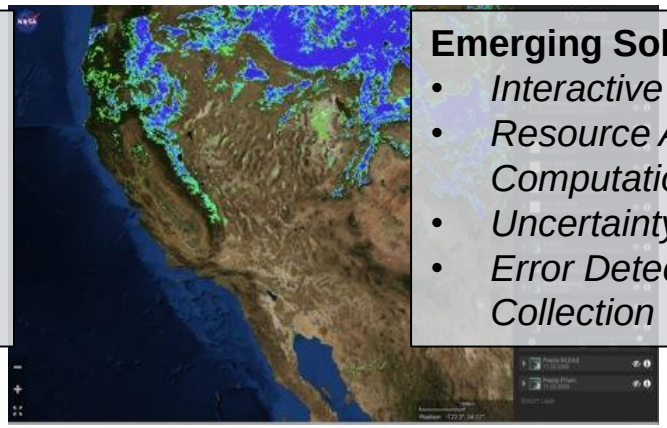
Emerging Solutions
- *Automated Machine Learning - Feature Extraction*
- *Intelligent Search*
- *Learning over time*
- *Integration of disparate data*

Technologies: Machine Learning, Deep Learning, Intelligent Search, Data Fusion, Interactive Visualization and Analytics

Agile MOS-GDS

Emerging Solutions
- *Anomaly Interpretation*
- *Dashboard for Time Series Data*
- *Time-Scalable Decision Support*
- *Operator Training*

Data Analytics and Decision Support



Emerging Solutions
- *Interactive Data Analytics*
- *Resource Analysis of Computational Workflows*
- *Uncertainty Quantification*
- *Error Detection in Data Collection*

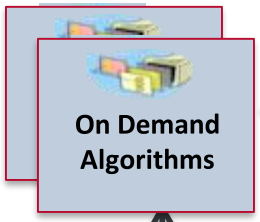
Credit: D. Crichton, R. Doyle

Future of Data Science at NASA

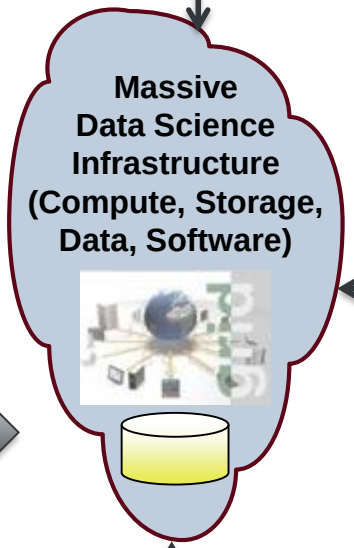
Shift Toward Data Analytics

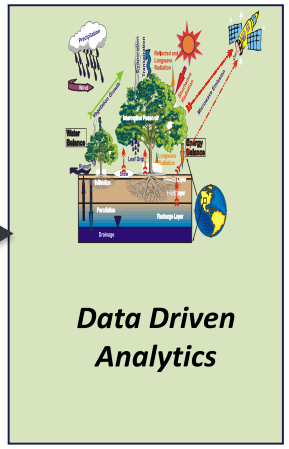
Data Stewardship

Today



Science Teams

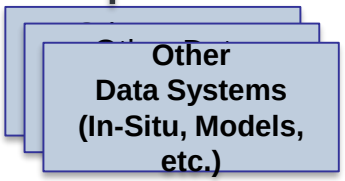




Research

Applications

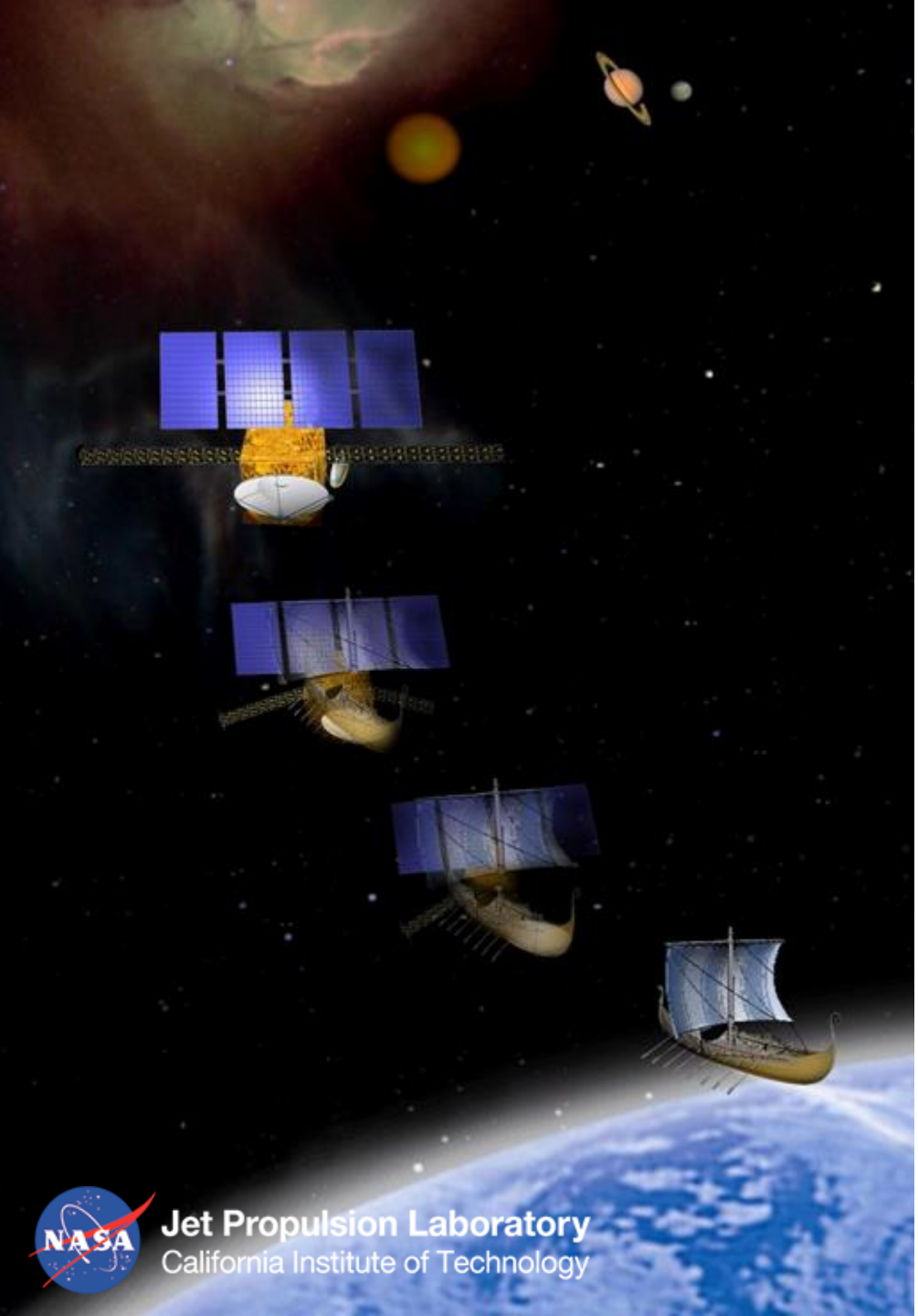
Decision
Support



Data Analytics

Future

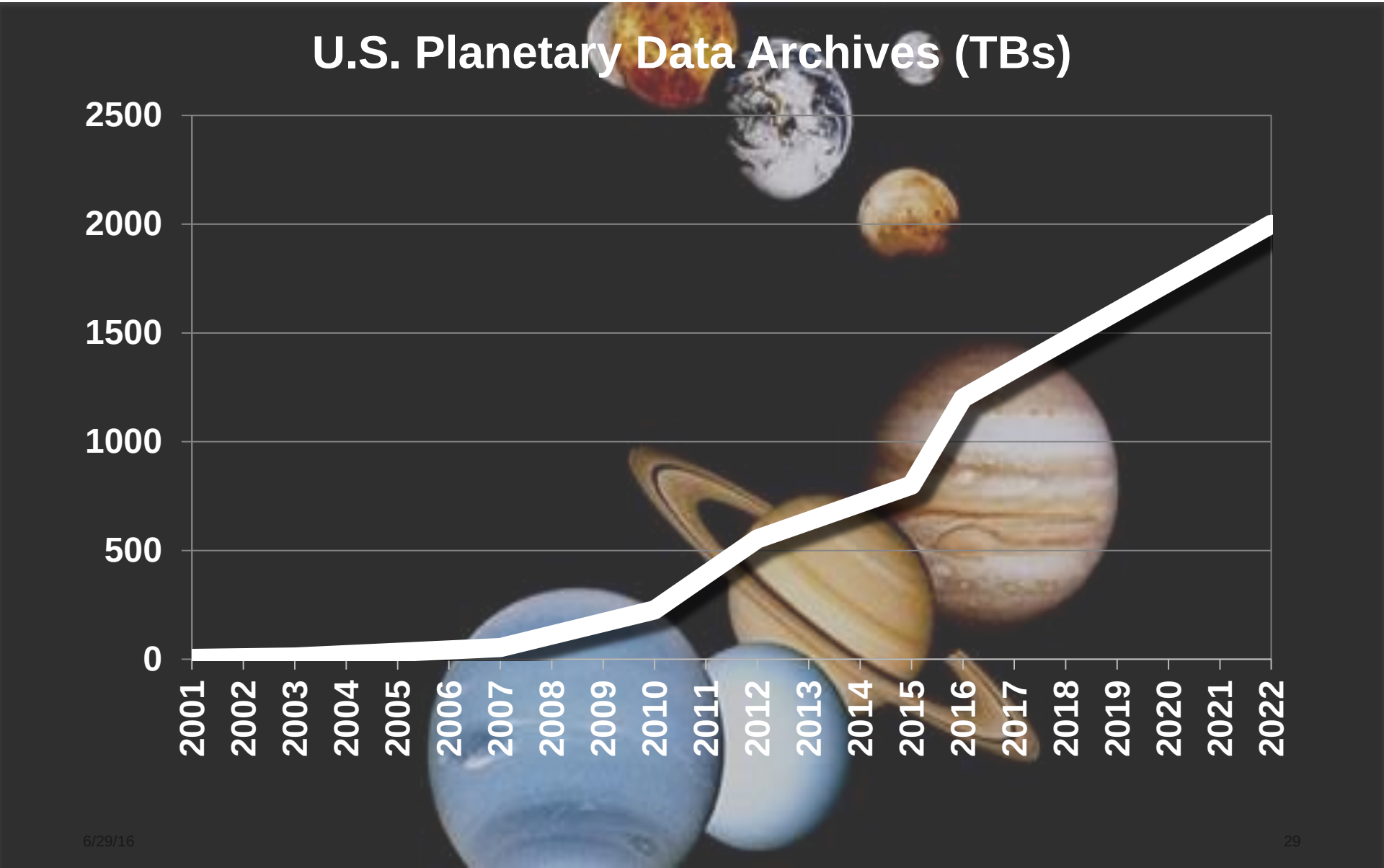
Credit: D. Crichton



Data-Driven Science Archives

The Planetary Data System

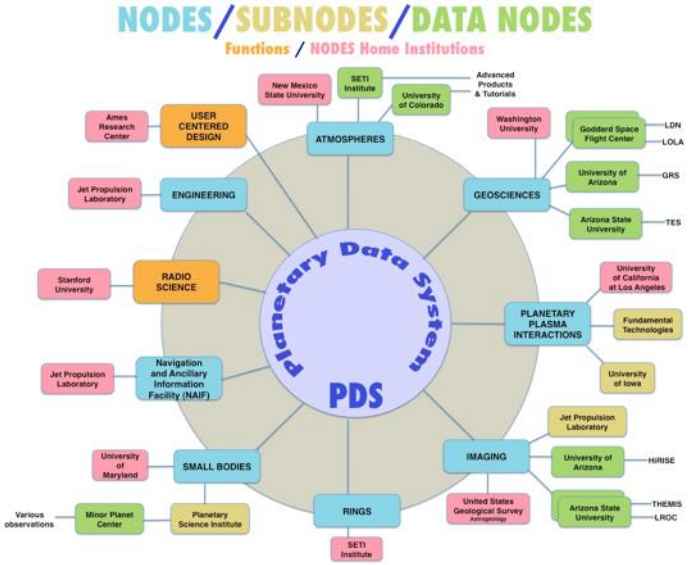
Growth of Planetary Data Archived from U.S. Solar System Research



Planetary Data System: An international data platform for Planetary Science Research

- Purpose: To collect, archive and make accessible digital data and documentation produced from NASA's exploration of the solar system from the 1960s to the present.

- Infrastructure: A highly distributed infrastructure with planetary science data repositories implemented at major government labs and academic institutions
  - System driven by a well defined planetary science information model
  - Over 1 PB of data
  - Movement towards international interoperability
  - Distributed federation of US nodes and international archives



30

Credit: D. Crichton

Big Data Challenges in Planetary Science

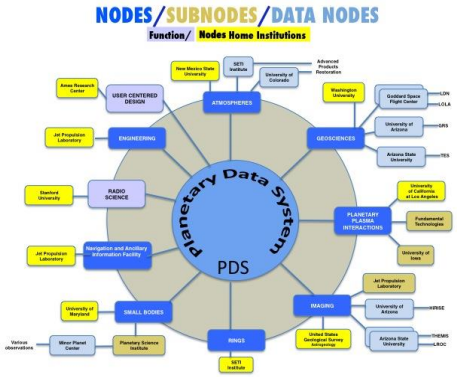
- Variety of planetary science disciplines, moving targets, and data
  - 4000 different types of data; 40M different observations
- Volume of data returned from missions including provenance
- Federation of disciplines and international interoperability

- These factors can affect choices in:
  - Data Consistency
  - Data Storage
  - Computation
  - Movement of Data
  - Data Discovery
  - Data Distribution

An Information Model-driven strategy to enable data management, discovery and analytics…

Credit: D. Crichton

Developing an Information-Model-Driven Strategy

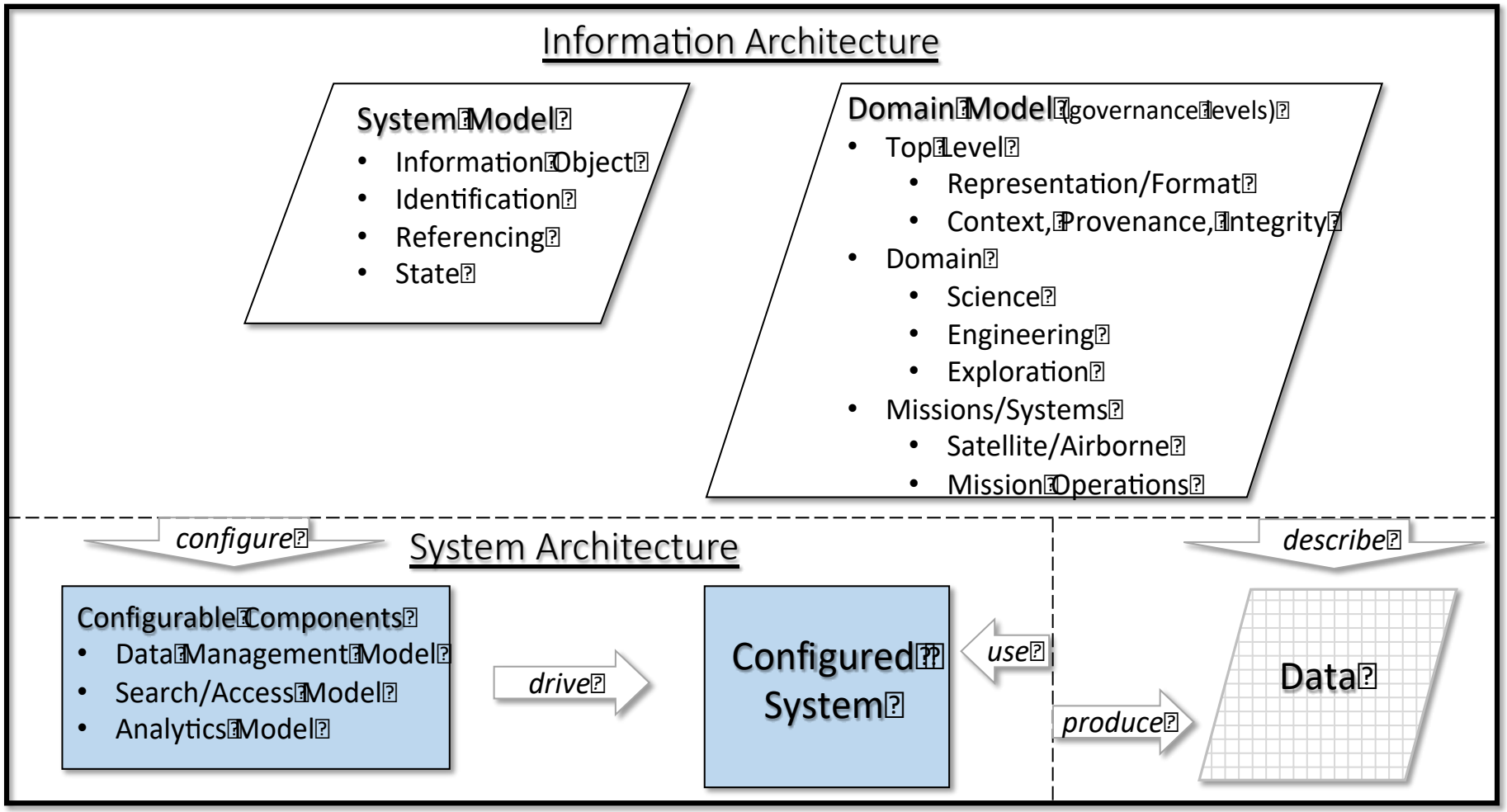


- An information model identifies and defines:
  - the elements to be processed
  - the entities that provide context for the elements
  - the relationships that provide meaning for the elements
- An information model increases productivity by
  - maximizing compatibility between systems (via reuse of common knowledge)
  - simplifying the process of design (via recurring patterns in the data)
  - promoting communication between individuals and teams working on the system (via common terminology and semantics)
- An information model is effective because
  - models are developed through extensive communication between domain experts, users, and model developers
  - models capture the points of view of users and experts who are most familiar with the domain
  - models serve as the basis for implementing the system software and processes

Credit: D. Crichton

Enabling a Model-Driven Data System

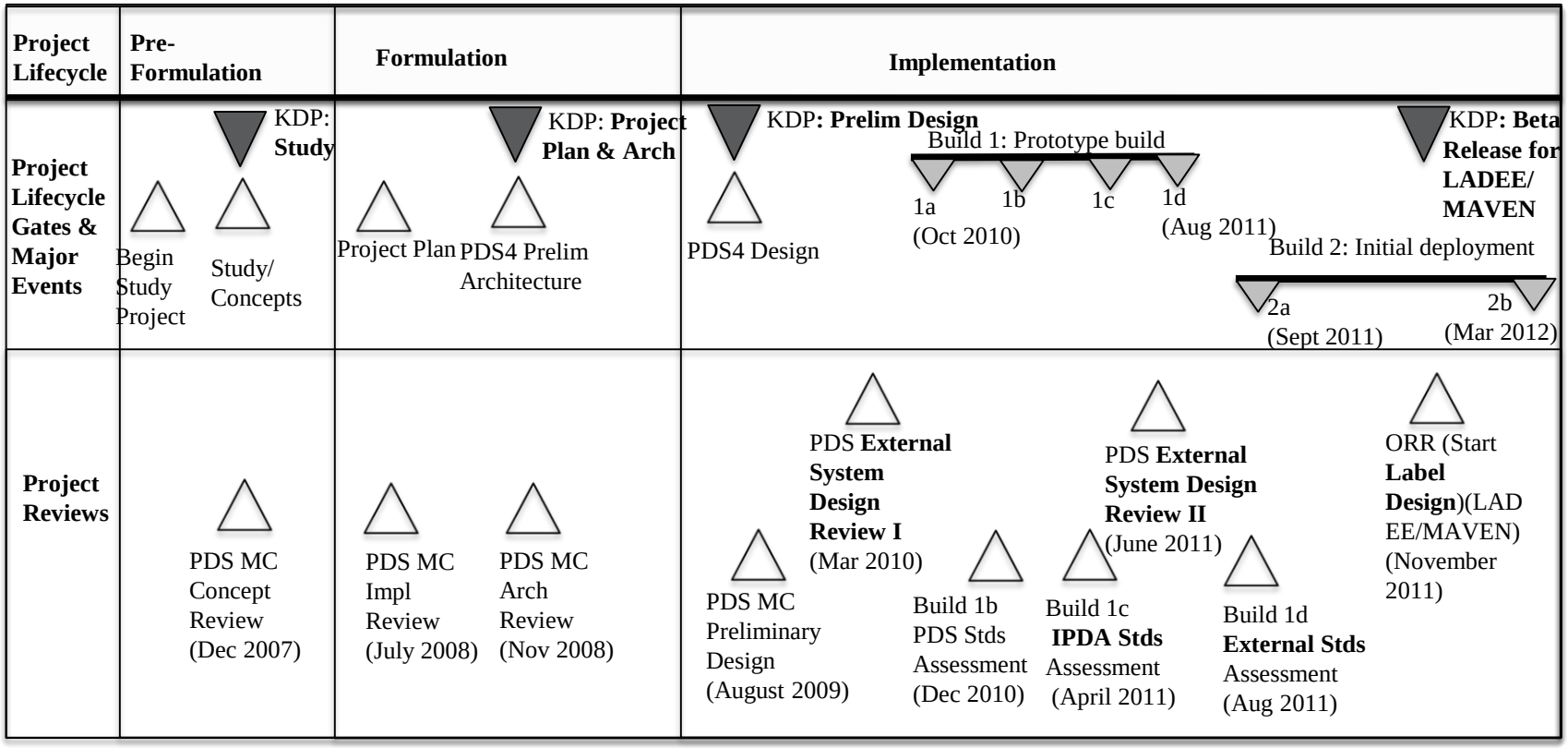
Information System Architecture



Crichton, D. Hughes, J.S. ; Hardman, S. ; Law, E. ; Beebe, R. ; Morgan, T.; Grayzeck, E.
A Scalable Planetary Science Information Architecture for Big Science Data.
IEEE 10th International Conference on e-Science, October 2014.

Project Implementation Approach

Regular builds and releases of the information model and software



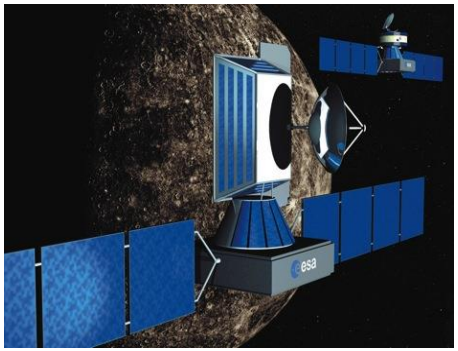
2014 NASA Award for the largest re-architecture in the history of the PDS (more than 30 years)

Credit: D. Crichton

Building a world-wide Data Science Platform for Planetary Science Research

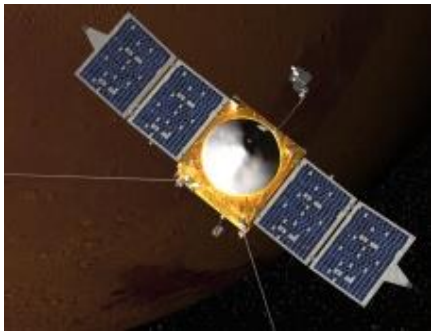
LADEE
(NASA)

InSight
(NASA)

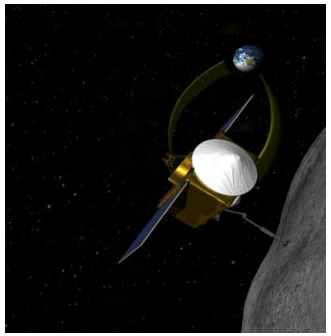


BepiColombo
(ESA/JAXA)

Mars 2020
(NASA)

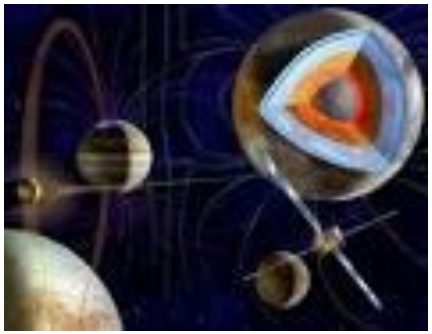


MAVEN
(NASA)



Osiris-REx
(NASA)

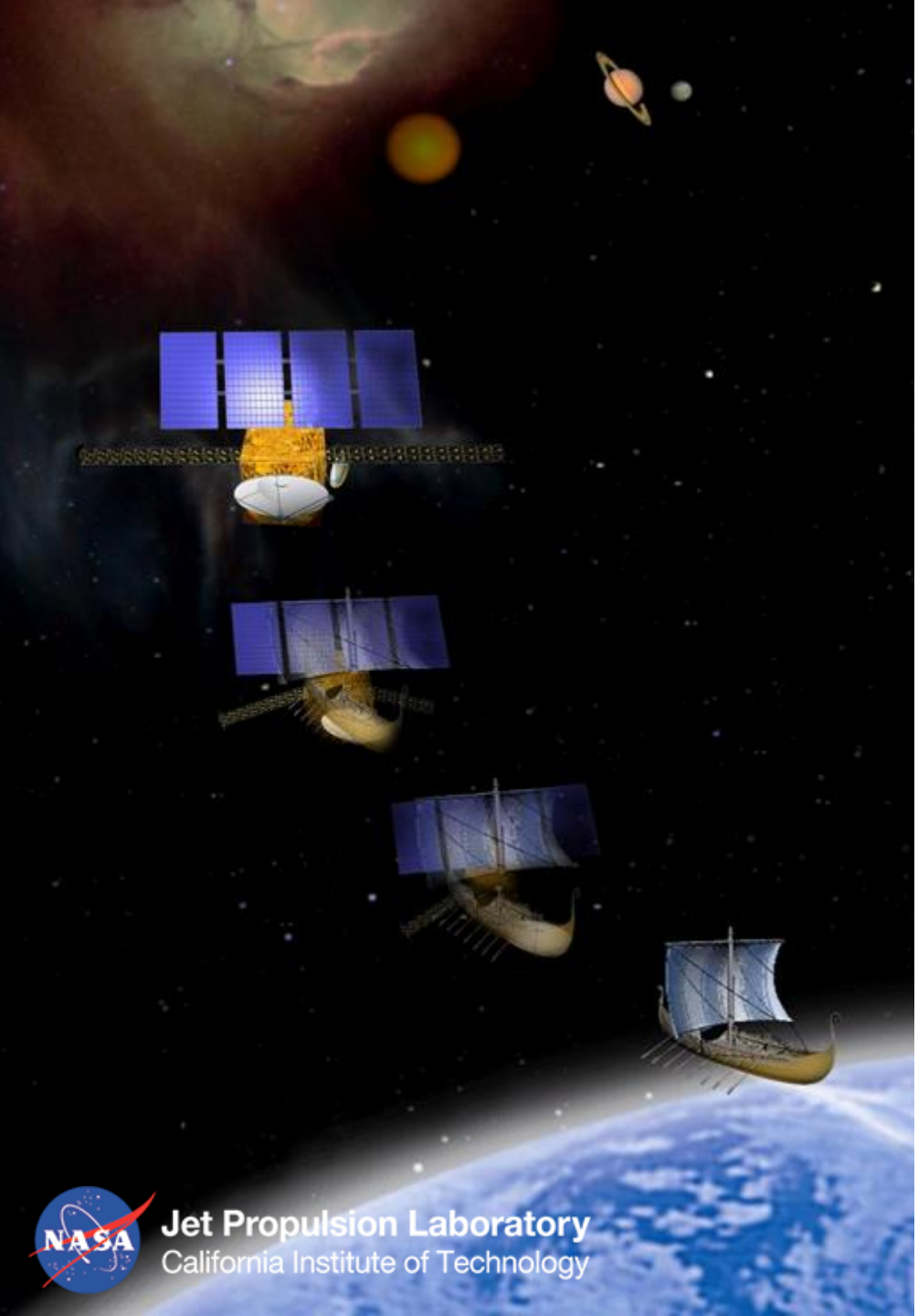
ExoMars
(ESA)



JUICE
(ESA)

…also Hayabussa-2, Chandrayaan-2

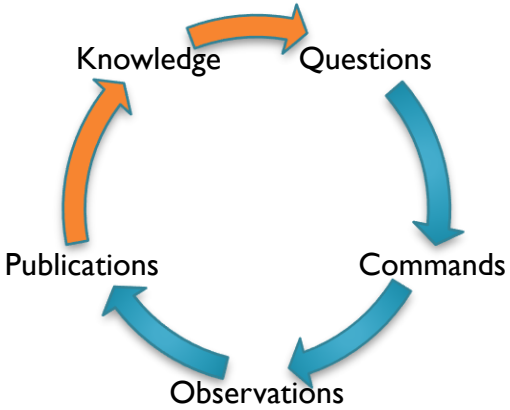
Endorsed by the **International Planetary Data Alliance** in July 2012 –
https://planetarydata.org/documents/steering-committee/ipda-endorsements-recommendations-and-actions



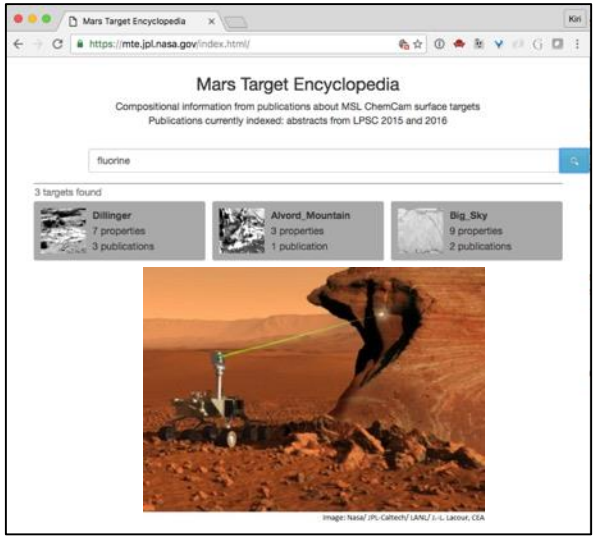
Looking Ahead

The Future of Analytics

Facilitate Science: Mars Target Encyclopedia

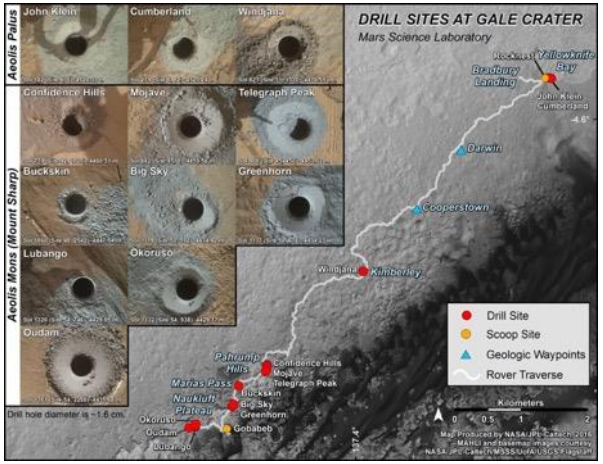


Help Scientists Do
What They Do Faster



- Connects data to (published) knowledge & maps

- Enables new searches:
  - What is known about Target X?
  - Scientific consensus?
  - Others like Target X?
  - What Targets are most unusual?
  - Show me publications that support statements



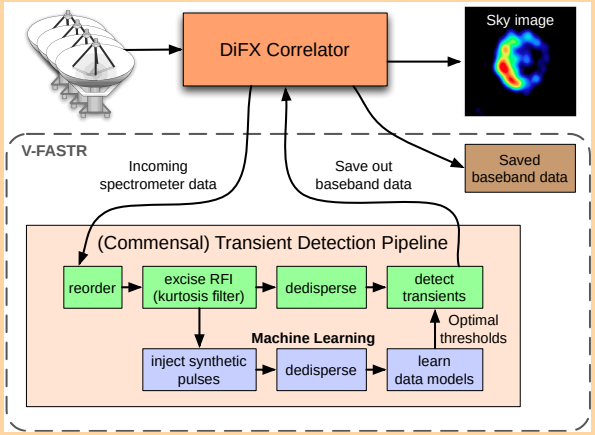
- Explore by geography, topic, publications, authors
- Speed scientific progress and exploration

Credit: K. Wagstaff et al

Data Triage, Analysis, and Understanding of Massive Data

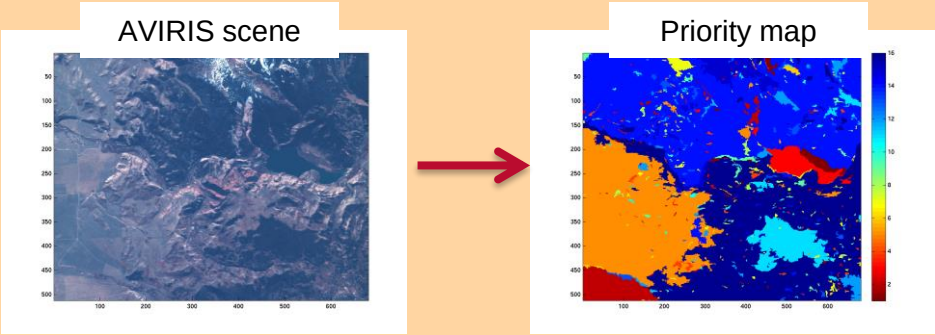
- Detection: fast identification of signals of interest (triage)

Radio astronomy:
V-FASTR realtime system at the VLBA



- Prioritization: use triage decisions to inform adaptive data compression

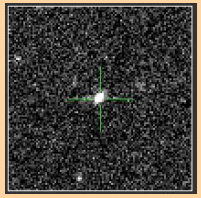
Earth science:
Onboard content-sensitive data compression

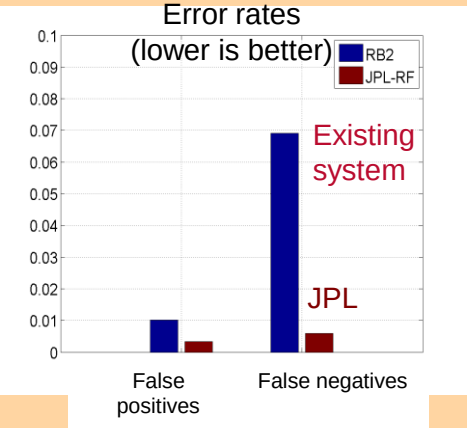


- Classification: online, real-time source type classification

Optical astronomy:
Reducing false positives for the Palomar Transient Factory

Real or spurious?

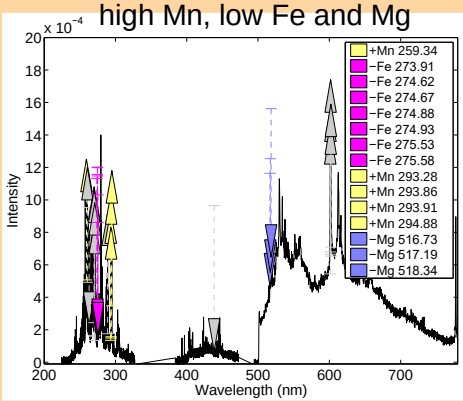




- Understanding: generate human-understandable explanations for decisions

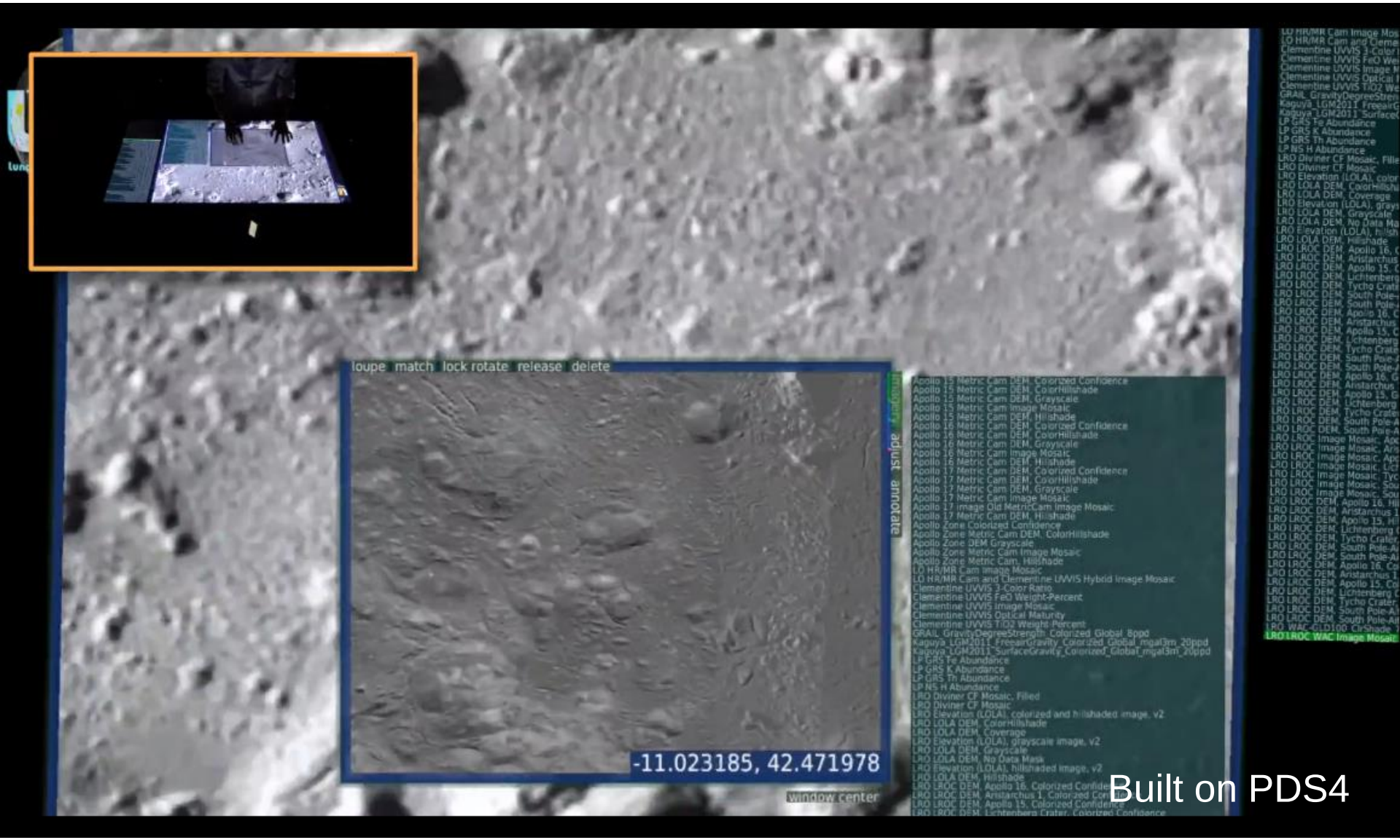
Planetary science:
Anomaly detection in ChemCam emission spectra from Mars, with content-sensitive "explanations" indicated with arrows (higher than expected vs. lower than expected)

Rhodochrosite explanations: high Mn, low Fe and Mg



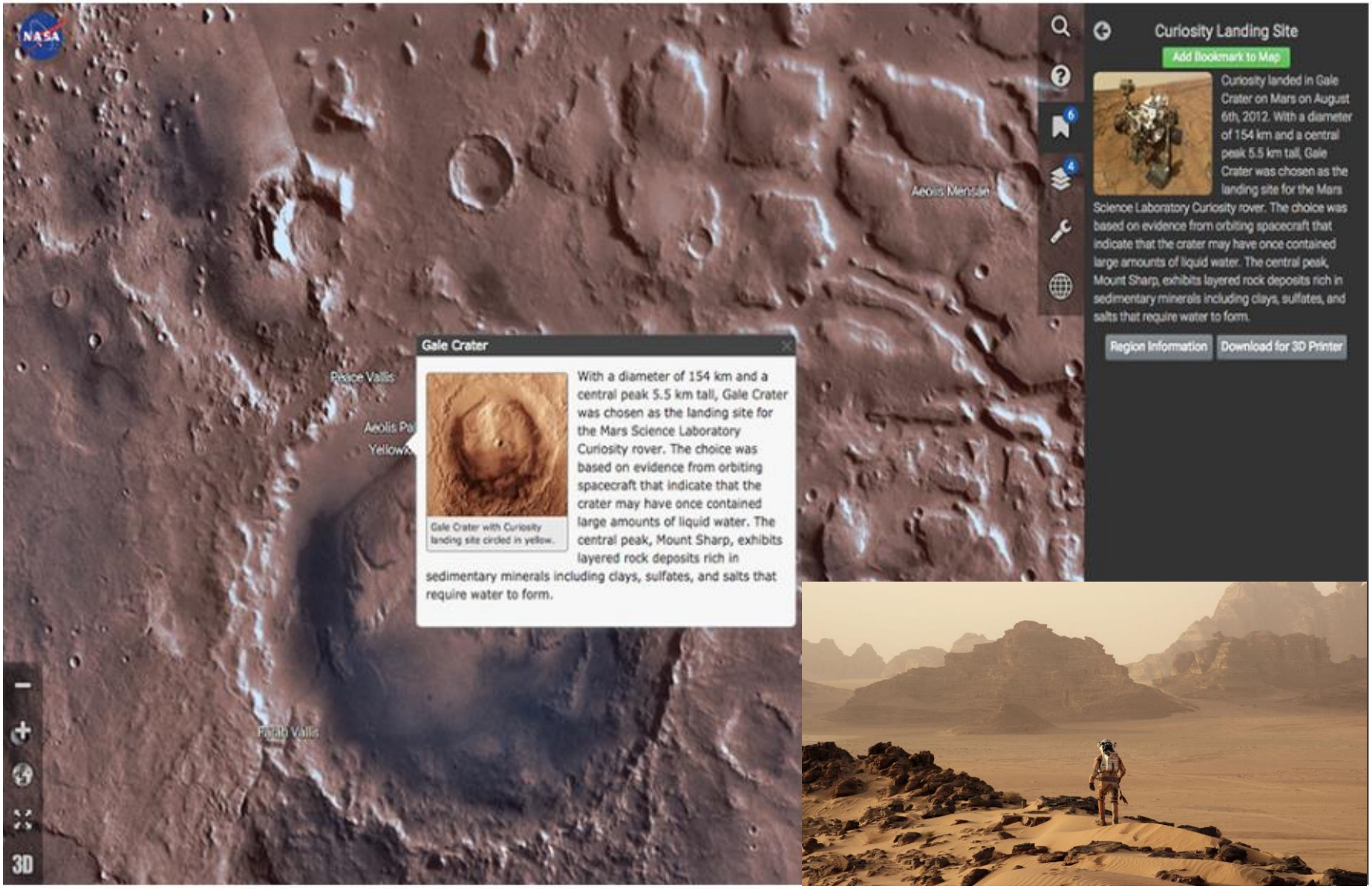
Credit: K. Wagstaff, U. Rebbapragada, D. Thompson, B. Tang, H. Xie

Lunar Trek
Exploring the Moon through Data Analytics



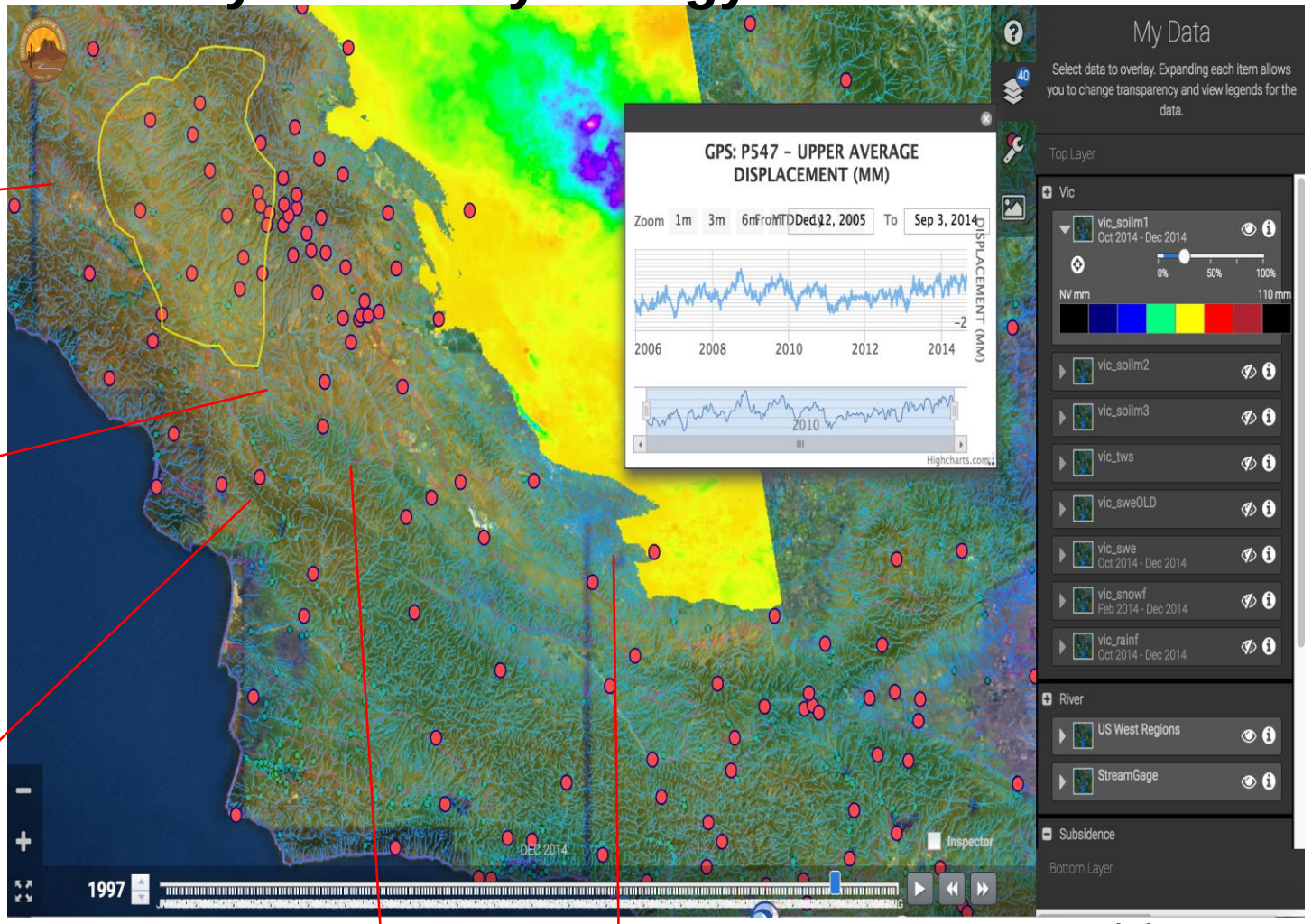
Credit: E. Law, S. Malhotra, G. Chang

Mars Trek:
The Google Earth of Mars



Credit: E. Law, S. Malhotra, G. Chang

WaterTrek:
Interactive Data Analytics for Hydrology



User Defined Polygon

GPS

In-Situ: Stream Gage Sensors

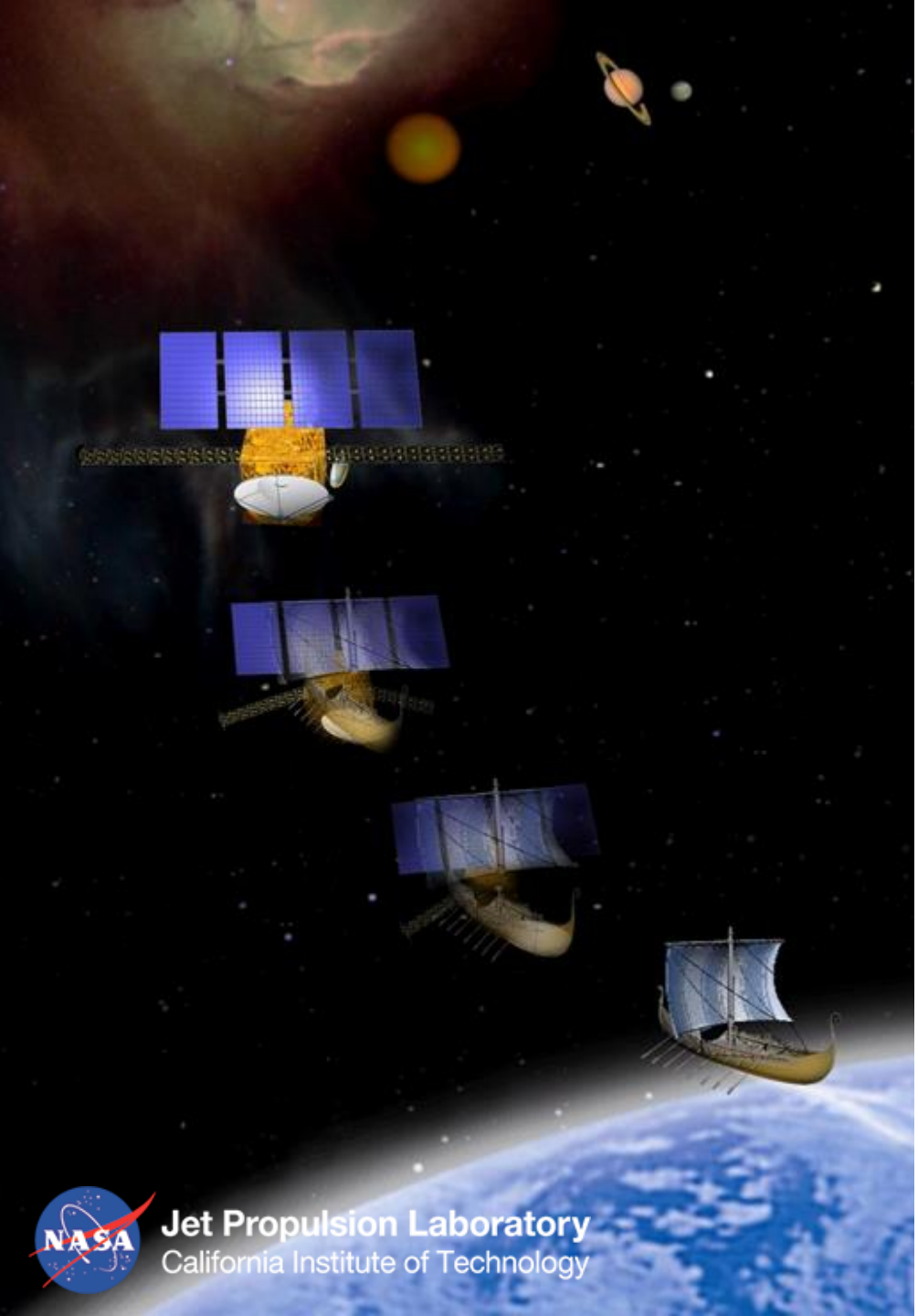
River Network

SAR derived Subsidence

Model Output: Soil Moisture

Fusing In-situ, Air-borne, Space-borne and model generated data using visualization and a big data analytics engine

Credit: S. Malhotra et al

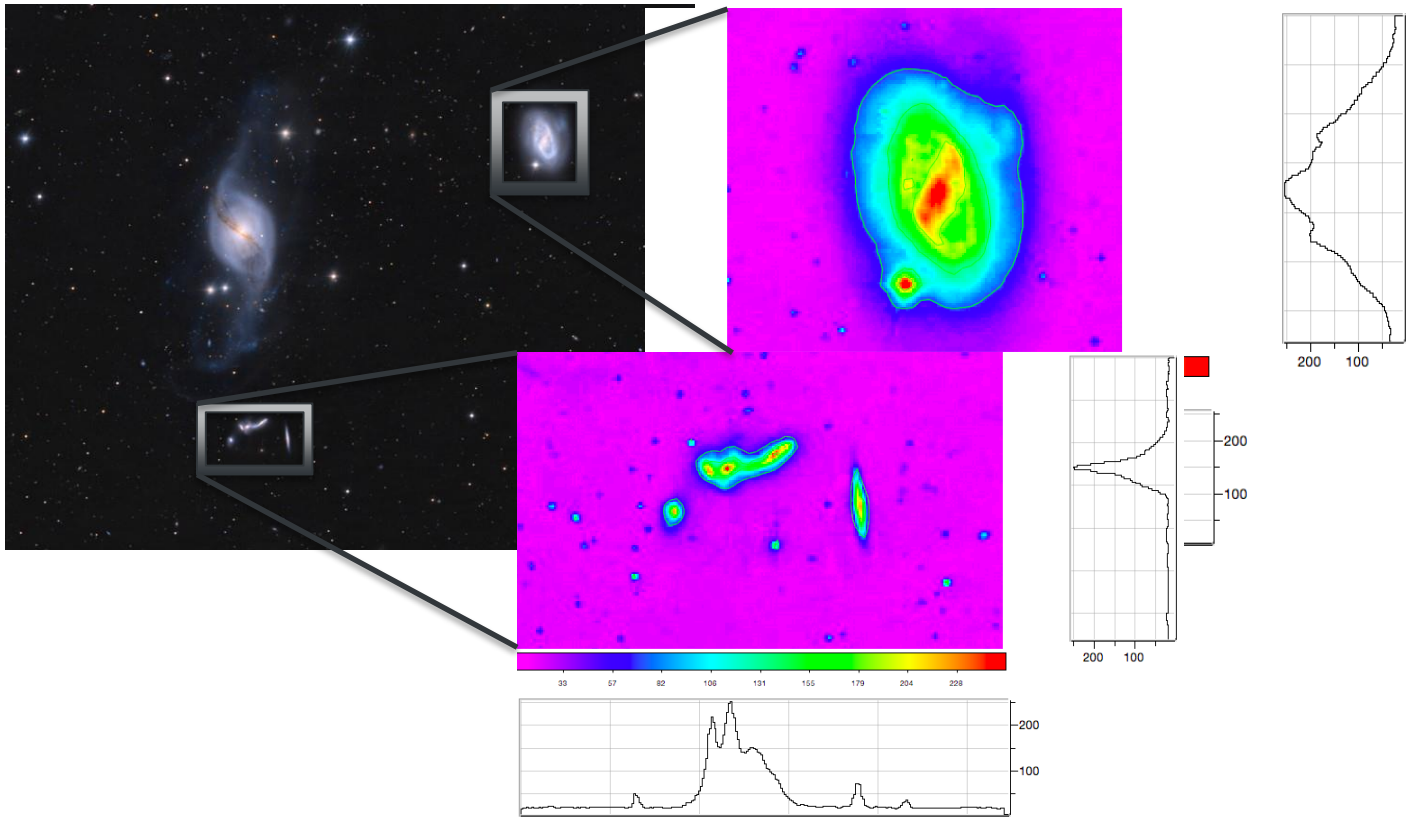


Looking Around

Methodology Transfer

Methodology Transfer

From Astrophysics…

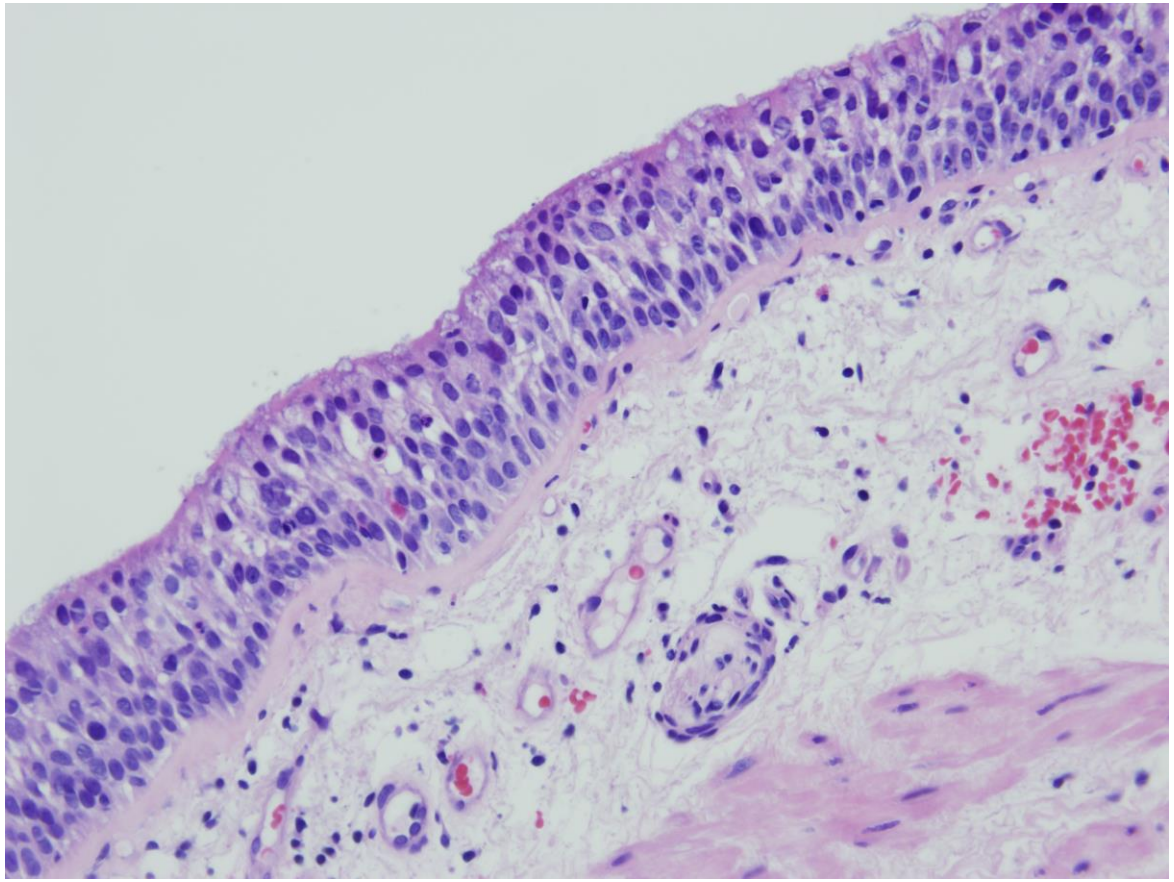


Description: Detecting objects from astronomical measurements by evaluating light measurements in pixels using intelligent software algorithms.

Image Credit: Catalina Sky Survey (CSS), of the Lunar and Planetary Laboratory, University of Arizona, and Catalina Realtime Transient Survey (CRTS), Center for Data-Driven Discovery, Caltech.

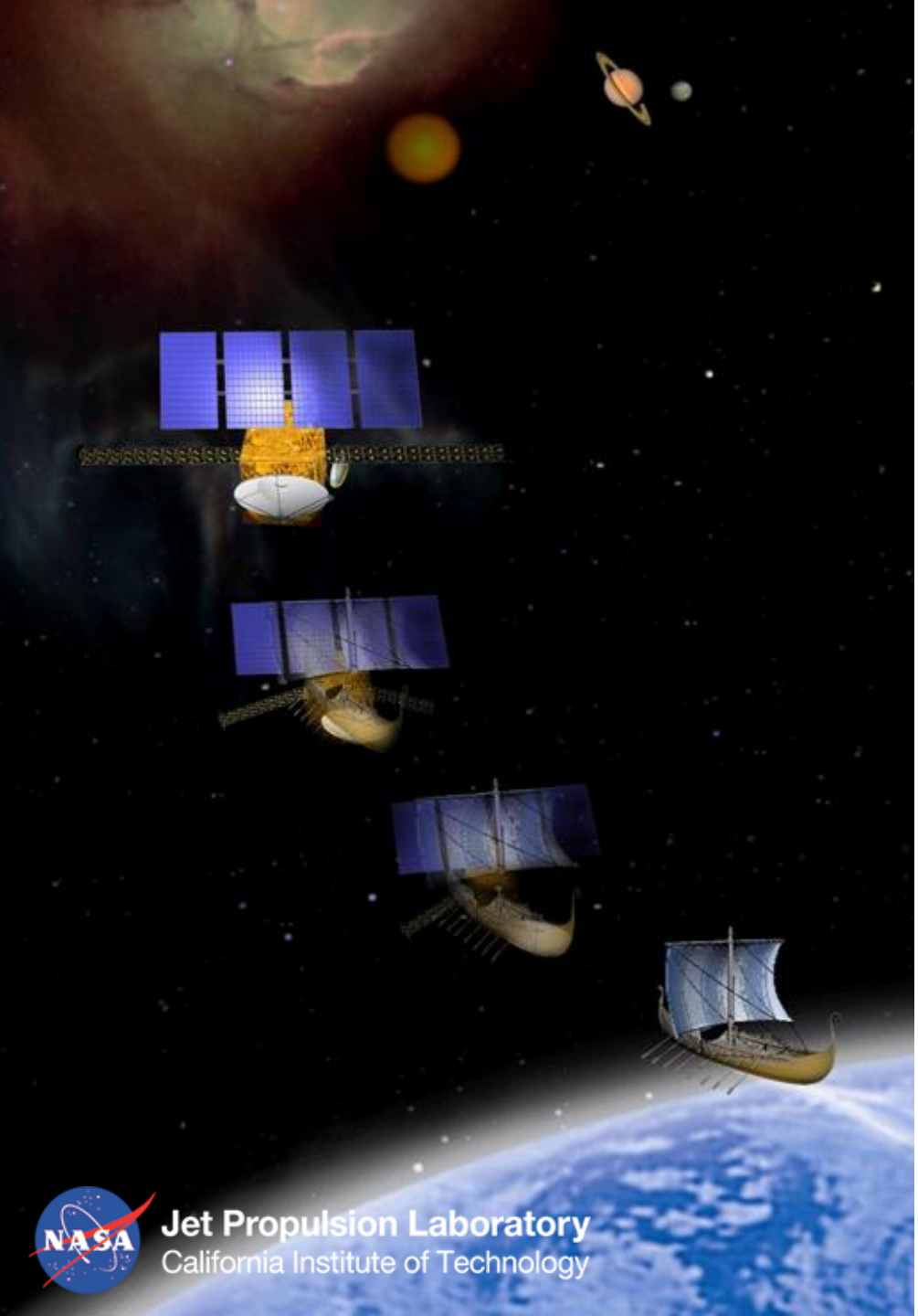
Methodology Transfer

…to Biomedicine



Description: Detecting objects from oncology images using intelligent software algorithms transferred to and from space science.

Image Credit: EDRN Lung Specimen Pathology image example, University of Colorado



Driving Forward

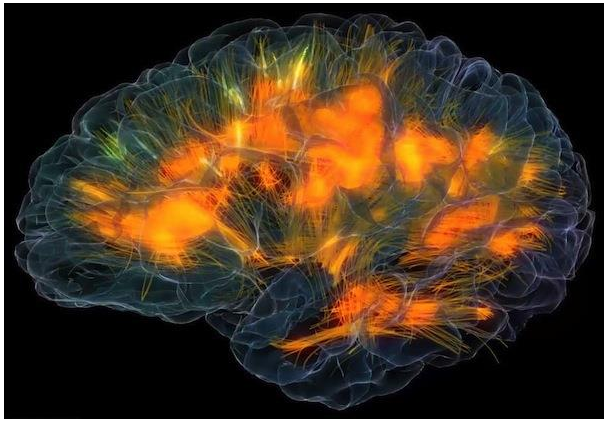
We're All in This Together

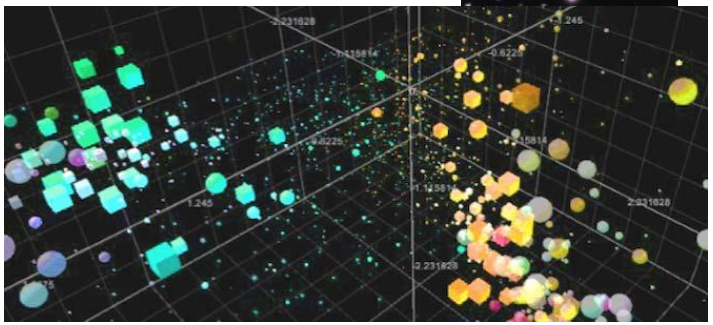
Caltech-JPL Partnership in Data Science and Technology

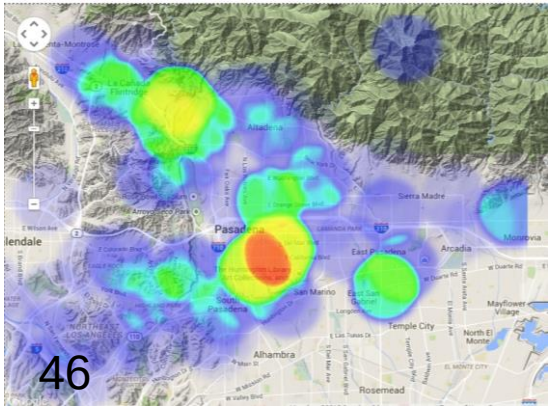
- Center for Data-Driven Discovery on campus/Center for Data Science and Technology at JPL
- From basic research to deployed systems ~10 collaborations
  - Leveraged funding from JPL to Caltech; from Caltech to JPL

CENTER FOR DATA-DRIVEN DISCOVERY

CDST





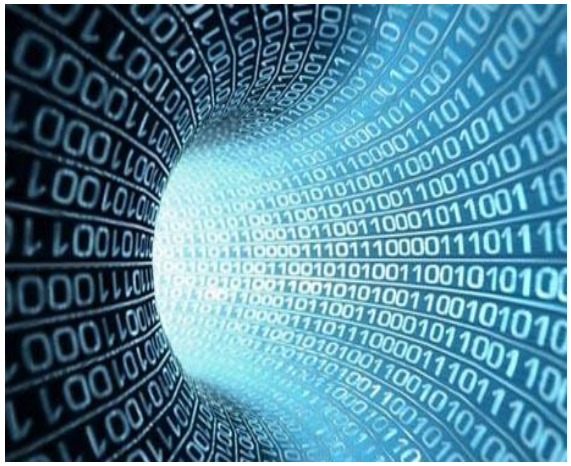


Opportunities for Collaboration

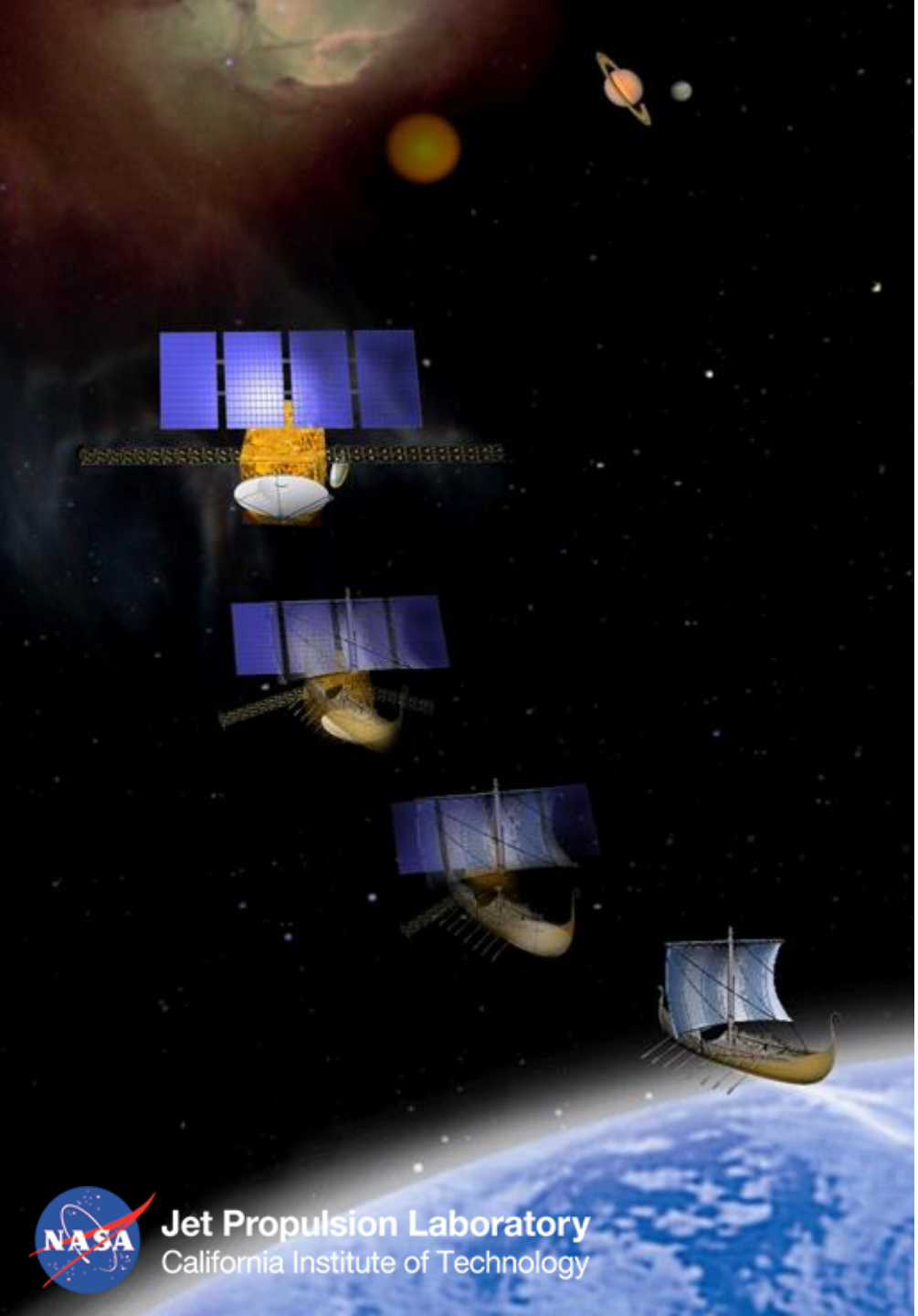
Partnerships

- Technology and Research Collaborations
  - Data-Driven Technologies
  - Novel Architectures
  - Technologies for the Data Lifecycle
  - Intersection between computer science, science, autonomy and robotics

What do we do with all this data?



This is looking like a black hole –
but wait, there's light at the end of the tunnel!


Jet Propulsion Laboratory
California Institute of Technology

Questions?