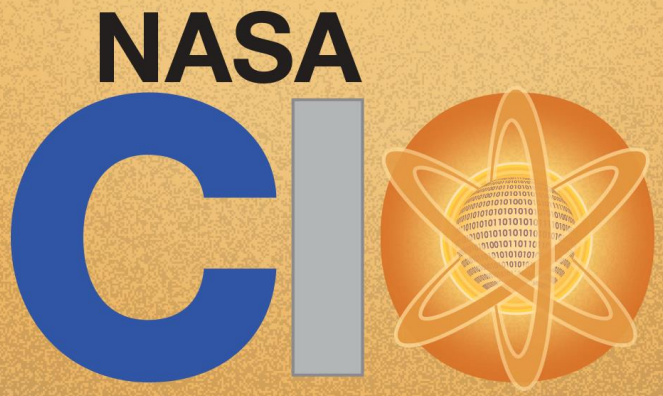
National Aeronautics and Space Administration

Data Science Exploration at NASA Langley Research Center

Jeremy Yagle
Information Management Branch, Office of the Chief Information Officer



NASA Research Centers









Research Centers

NASA Flight Centers



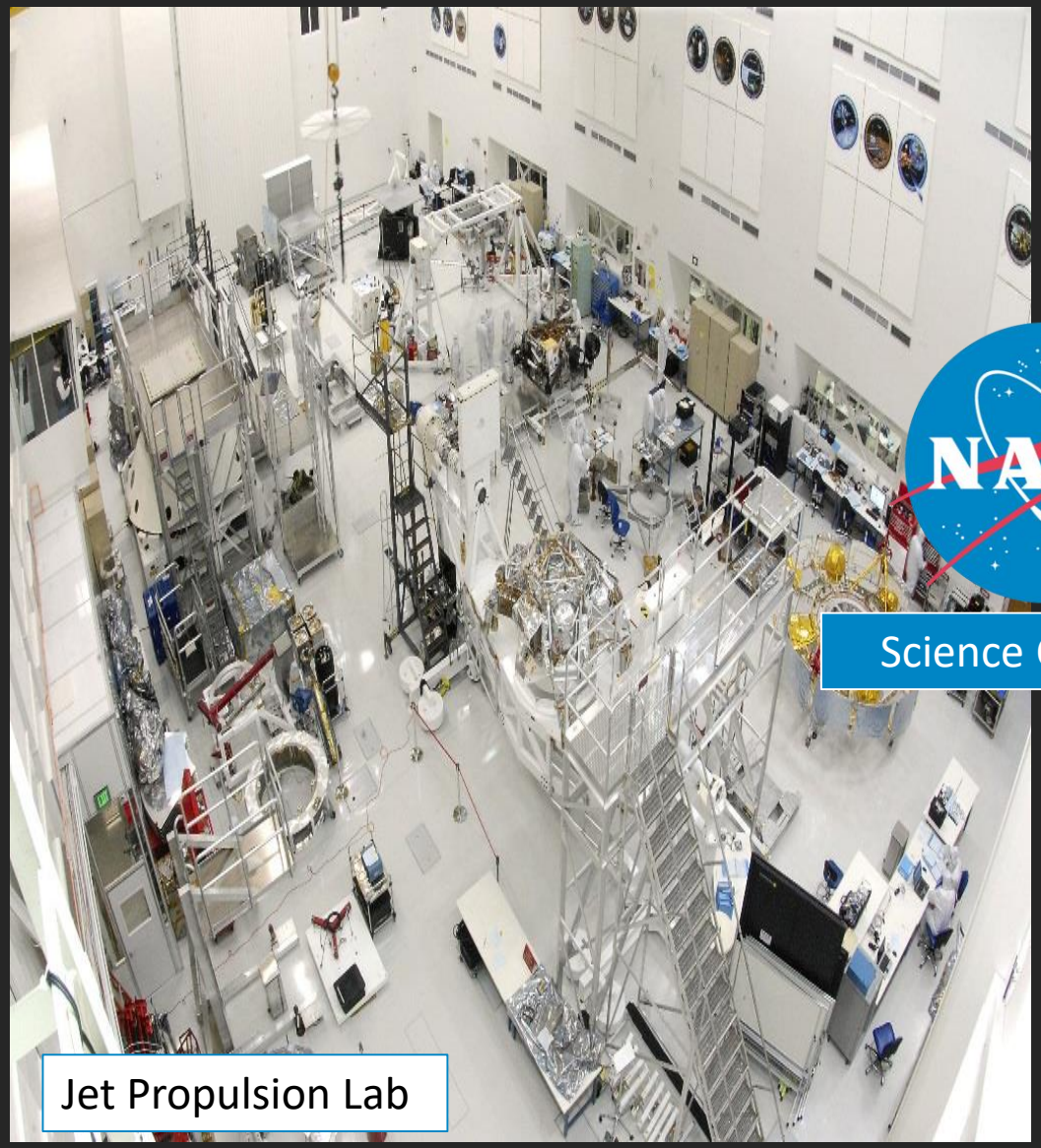






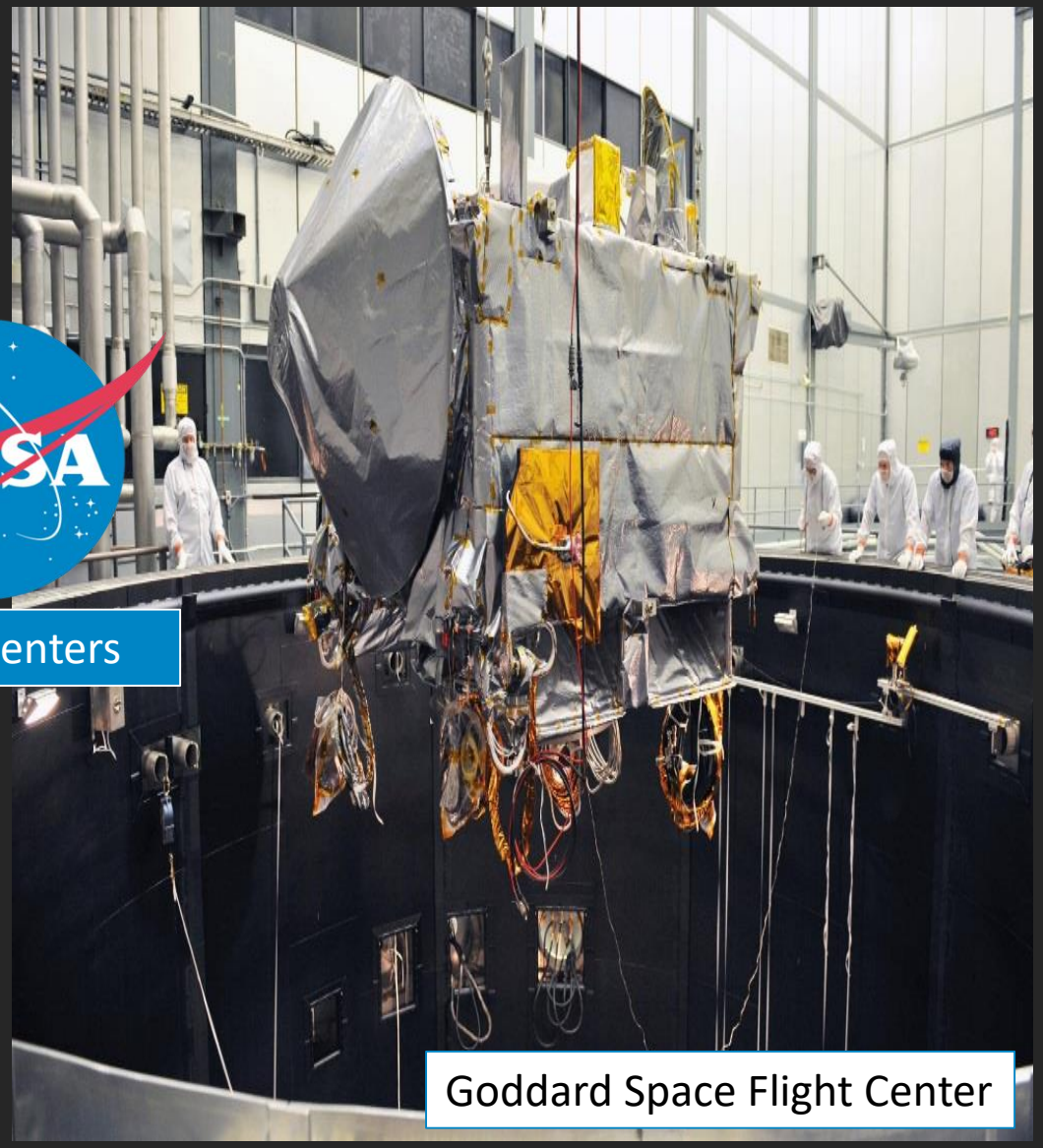
Flight Centers

NASA Science Centers



Jet Propulsion Lab

Science Centers



Goddard Space Flight Center

NASA Langley Research Center (LaRC)

Aerosciences



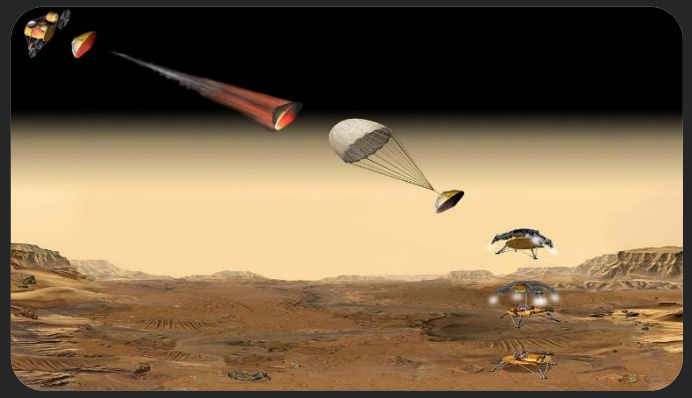
Systems Analysis & Concepts



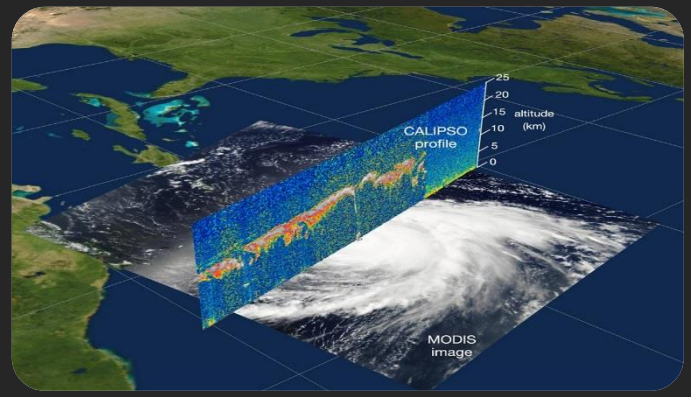
Intelligent Flight Systems



NASA Langley is a research, science, technology and development center that provides game-changing innovations to enable NASA to make significant contributions to the nation



Entry, Descent & Landing



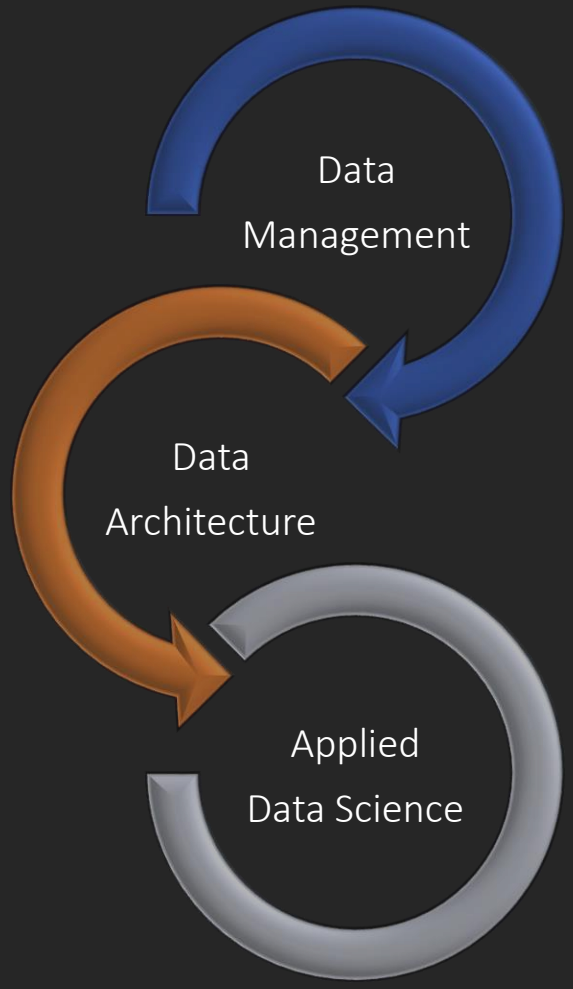
Atmospheric Characterization



Advanced Materials & Structural Systems

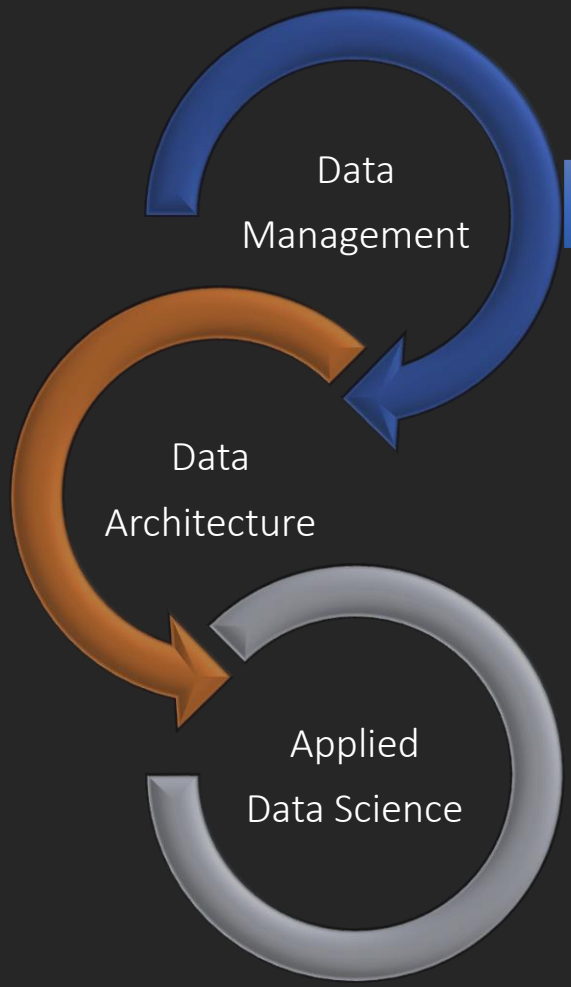
Supporting the NASA Mission: Data Science and OCIO

At LaRC, our data science efforts within the Office of the Chief Information Officer have focused on a broad-spectrum approach in key areas:



Encouraging an enterprise-level, platform-centric strategy of application development and deployment

Supporting the NASA Mission: Data Science and OCIO



Identify → Aggregate → Catalog → Share

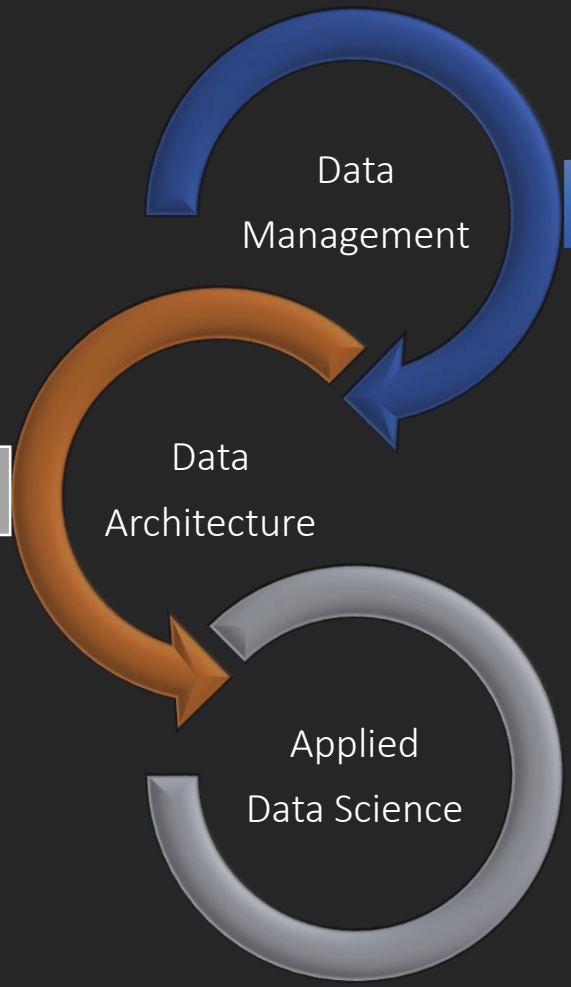
Ensuring data is collected and maintained in accordance with policy and best practices

Encouraging an enterprise-level, platform-centric strategy of application development and deployment

Supporting the NASA Mission: Data Science and OCIO

Supporting Mission partners by providing capabilities to organize and standardize

Engage → Distribute → Enhance → Steward

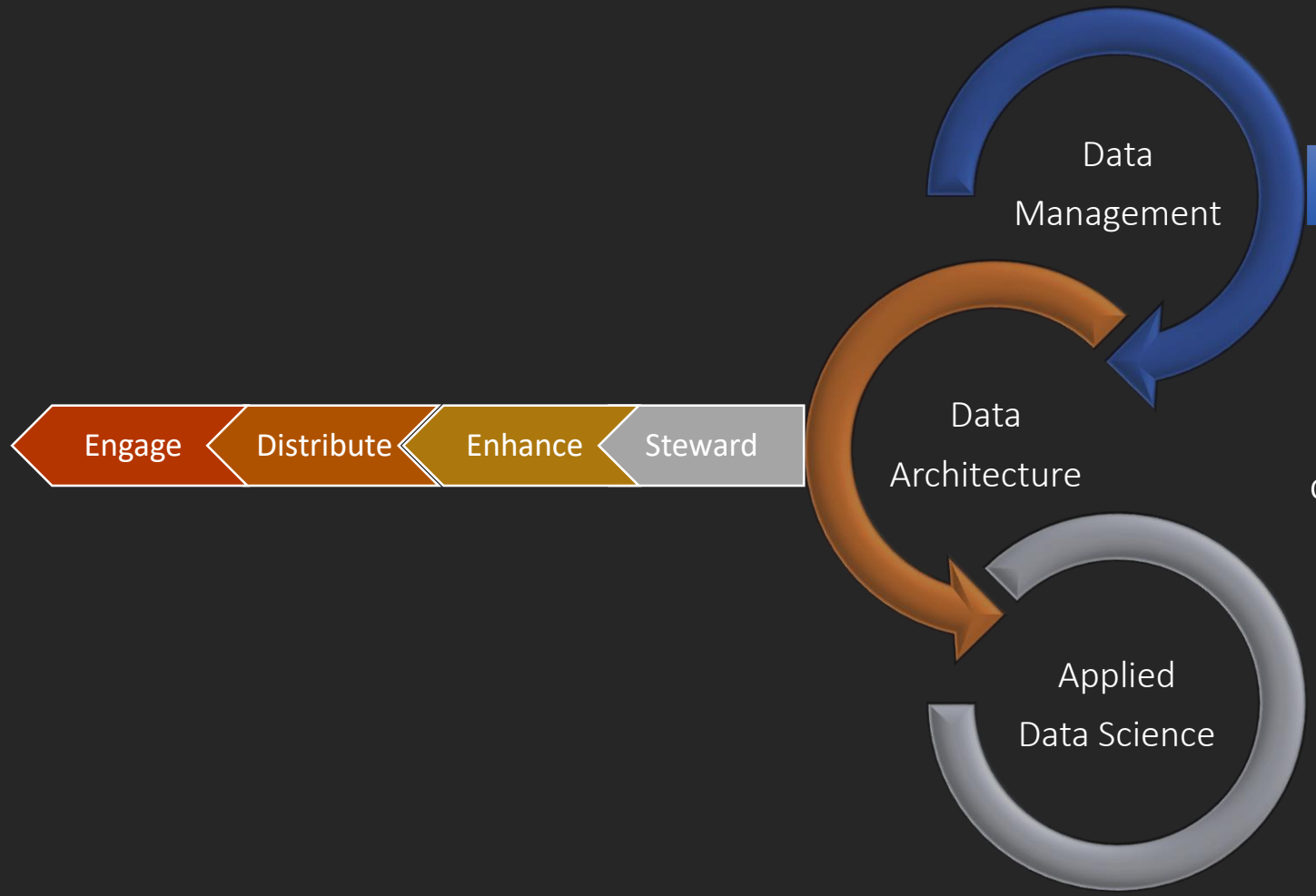


Identify → Aggregate → Catalog → Share

Identifying opportunities to enhance research and business processes:

- Materials Science
- Control Theory & Aerodynamics
- Contingency Management
- Climate Science
- Financial & Project Management

Supporting the NASA Mission: Data Science and OCIO



Over 6 years of use case development in a variety of aerospace domains have allowed us to identify a suite of common methods:

- Data Visualization
- Anomaly Analysis
- Precursor Detection
- Text Analytics
- Natural Language Processing
- Emergent Behavior
- Active Learning

Supporting the NASA Mission: Data Science and OCIO

Ensuring data is collected and maintained in accordance with policy and best practices

Identify → Aggregate → Catalog → Share

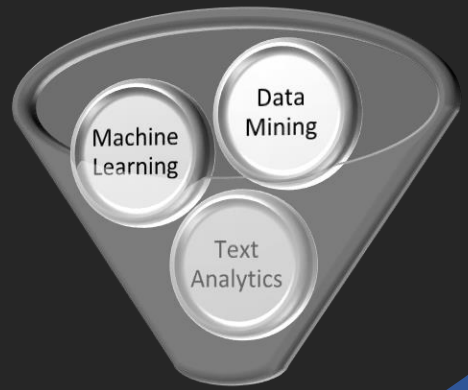
Data Management

Supporting Mission partners by providing capabilities to organize and standardize

Engage → Distribute → Enhance → Steward

Data Architecture

Over 6 years of use case development in a variety of aerospace domains



Applied Data Science

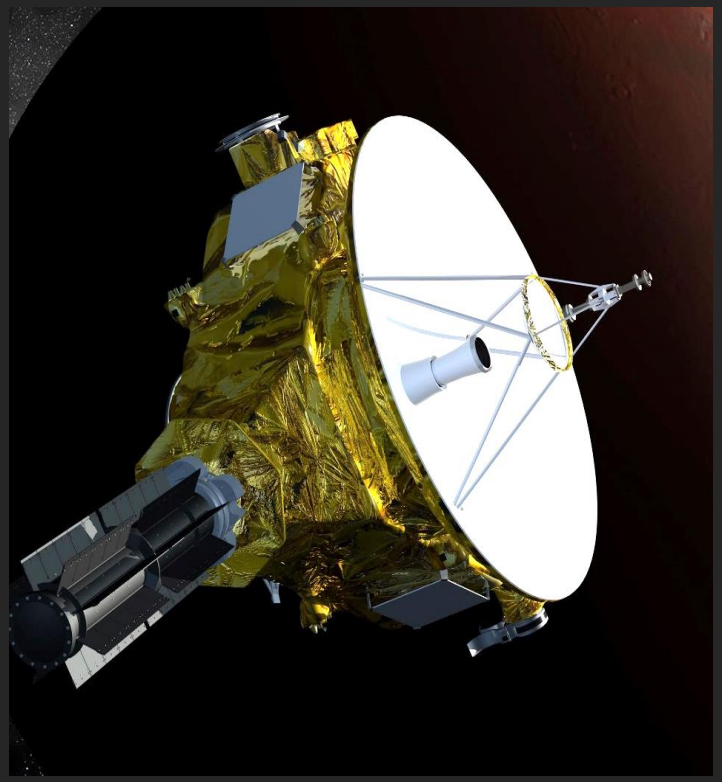
Encouraging an enterprise-level, platform-centric strategy of application development and deployment

To accomplish NASA's Mission, we need:

Large-scale capabilities with substantial power



Advanced technologies and innovative approaches



Quantification of uncertainty in models

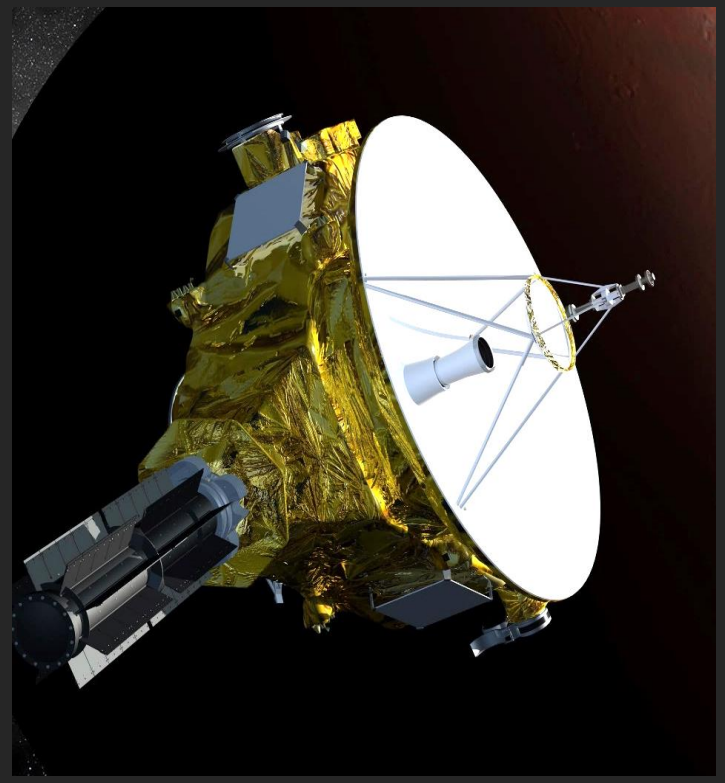


Likewise, for effective applications of data science & machine learning…

Some data requires a large-scale capability with substantial power…



…while other *data* requires advanced technologies and innovative approaches...

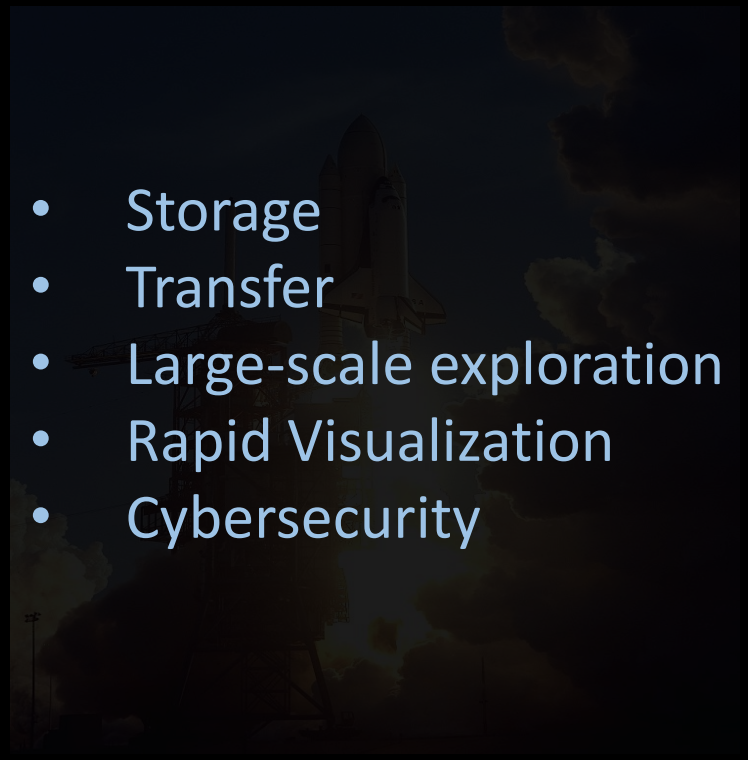


…and all *data* must support quantification of uncertainty in models.



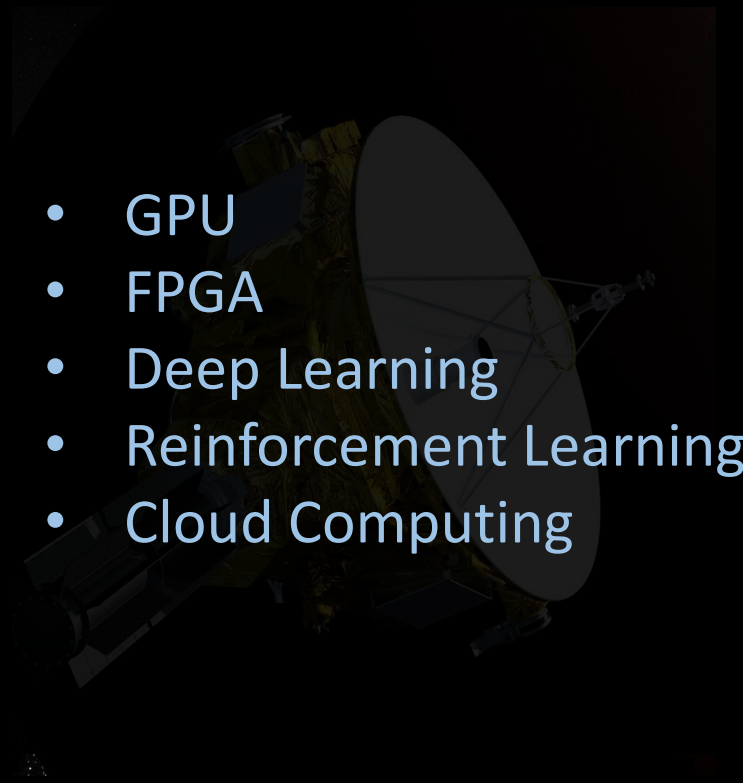
Likewise, for effective applications of data science & machine learning…

Large-scale capability with substantial power



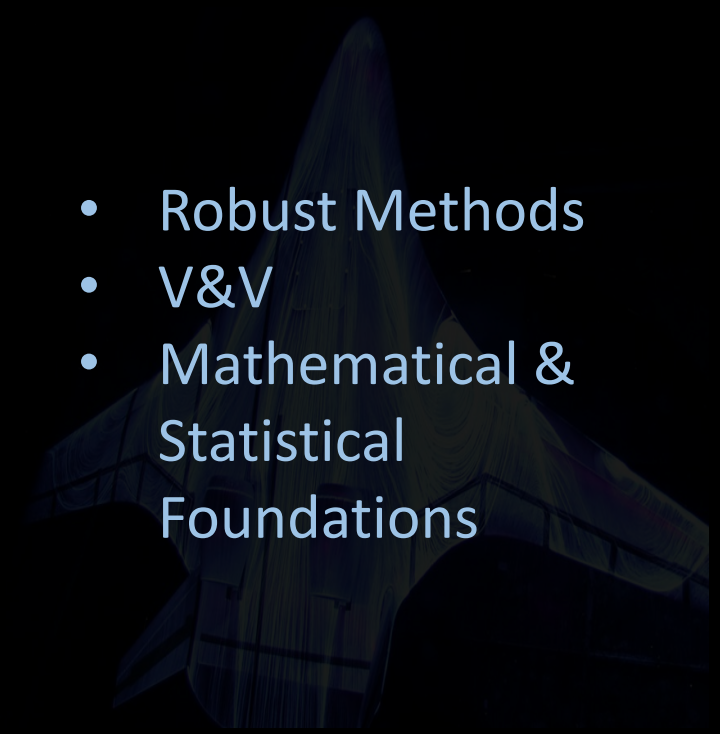
- Storage
- Transfer
- Large-scale exploration
- Rapid Visualization
- Cybersecurity

Advanced technologies and innovative approaches



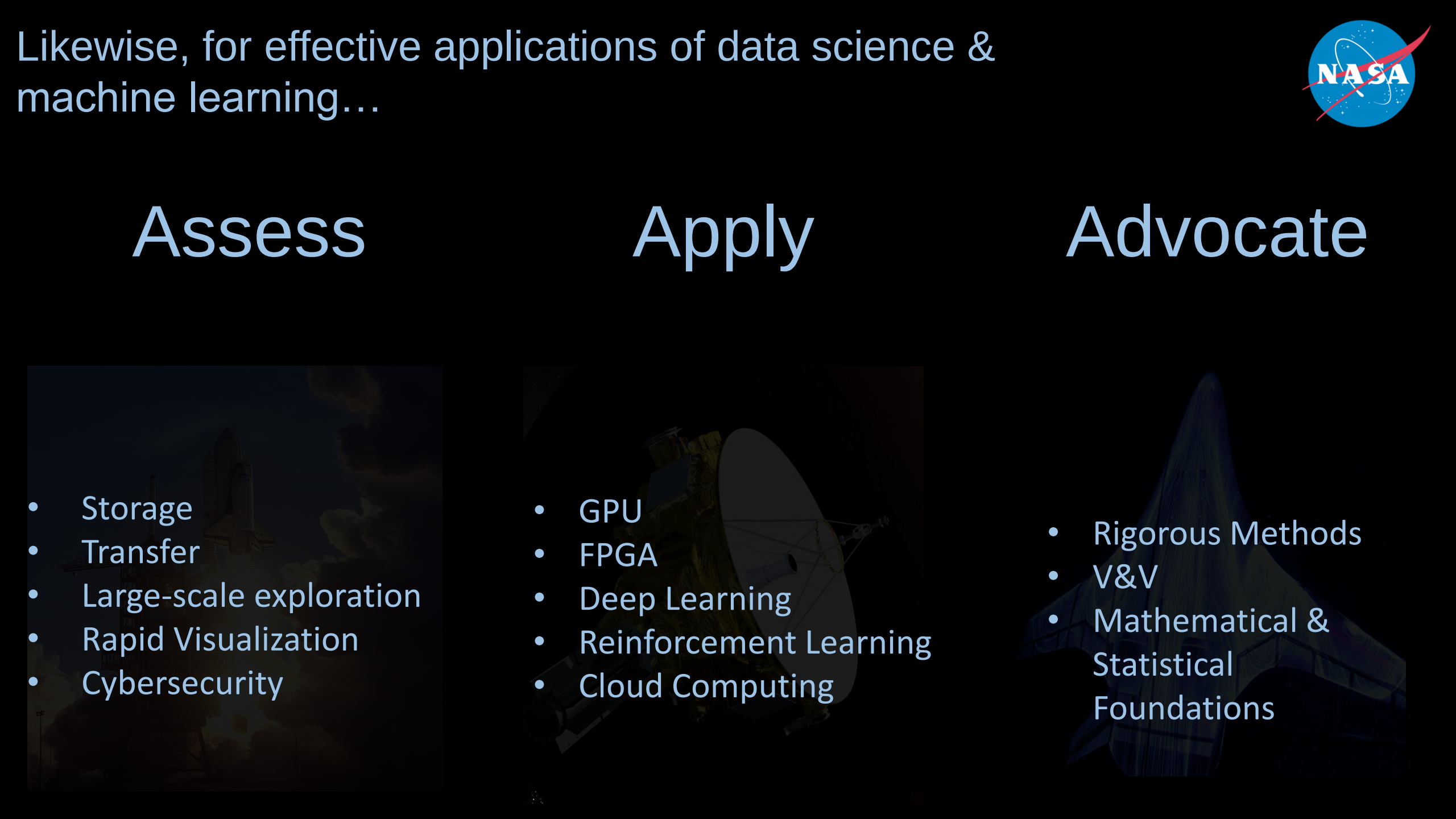
- GPU
- FPGA
- Deep Learning
- Reinforcement Learning
- Cloud Computing

Quantification of uncertainty in models



- Robust Methods
- V&V
- Mathematical & Statistical Foundations

Likewise, for effective applications of data science & machine learning…



Assess

- Storage
- Transfer
- Large-scale exploration
- Rapid Visualization
- Cybersecurity

Apply

- GPU
- FPGA
- Deep Learning
- Reinforcement Learning
- Cloud Computing

Advocate

- Rigorous Methods
- V&V
- Mathematical & Statistical Foundations

Assess, Apply, Advocate

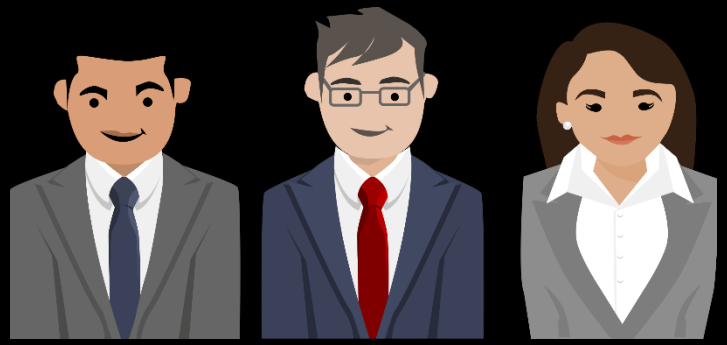
Data Science

For all technologies and applications, consider potential impacts.



CYBERSECURITY

Coordinate with:
- *IT Security*
- *IT Acquisition Teams*
- *Legal*



SOCIAL

Conduct outreach to understand:
- *Current challenges*
- *Long-Term goals*
- *Possible partners*



EASE OF USE

Identify solutions that are:
- *Accessible*
- *Intuitive*
- *Useful*



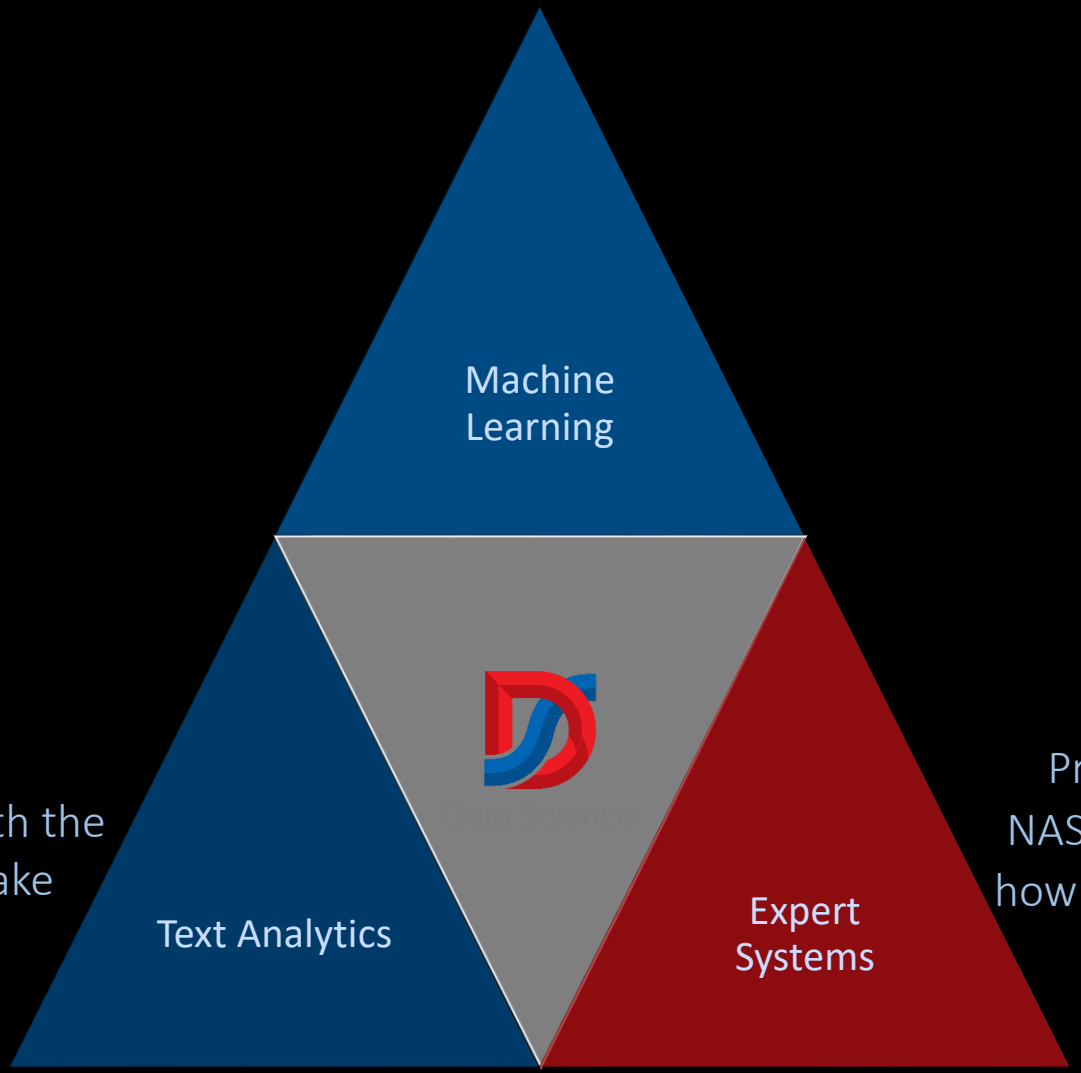
COST

Consider & Measure:
- *Return on Investment*
- *Open-source alternatives*
- *Possible partners*

Identify Capabilities

Data Science

Data-enabled methods to help experts make new discoveries



Tools for fast analysis of technical documents, with the ability to help experts make connections

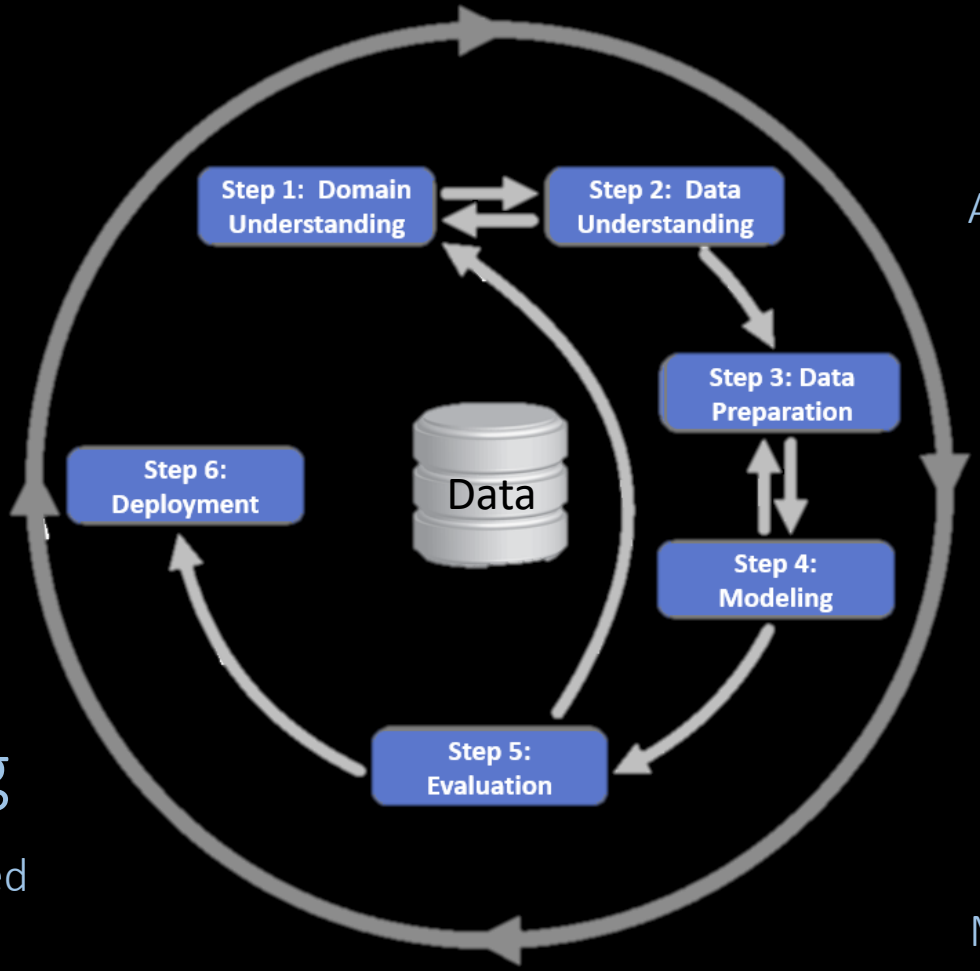
Proof of Concept systems in NASA domains, to understand how we can leverage cognitive technologies

Define the Process

Data Science

Knowledge Discovery in Databases

Extracting previously unknown information from data



CRISP-DM

Data Mining

Application of specific algorithms for detecting patterns and relationships in data

Machine Learning

Classification or prediction, based on properties learned from training data

Reinforcement Learning

Machine learning that makes the optimal choice in a given situation, based on rewards

Data Science in NASA Domains

Data Science

Step 1: Domain Understanding

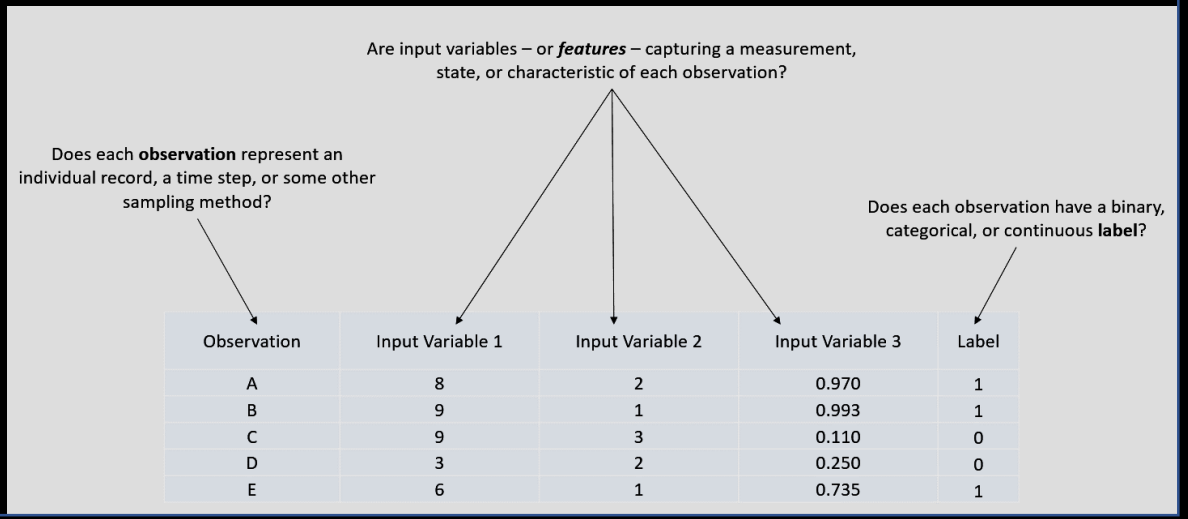
Understand project objectives and requirements

- Background
- Current State vs. Desired State
- Data Collection and Analysis Methods
- Technical Objectives
- Challenges and Risks
- Requirements and Scoping
  - Proposed Techniques
  - Literature Review

Step 2: Data Understanding

Initial exploration and identification of methods

- Assess curation and metadata
- Exploratory data analysis
- Determine what data science methods are viable, given data size and characteristics



Observation	Input Variable 1	Input Variable 2	Input Variable 3	Label
A	8	2	0.970	1
B	9	1	0.993	1
C	9	3	0.110	0
D	3	2	0.250	0
E	6	1	0.735	1

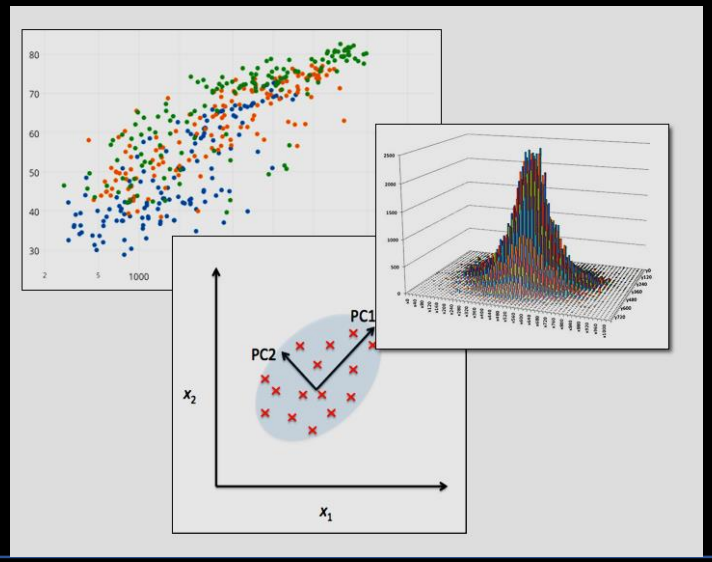
Data Science in NASA Domains

Data Science

Step 3: Data Preparation

Normalization, transformation, dimensionality reduction

- "Screen and Clean"
- Normalize and transform to ensure all data is subjected to an equivalent analysis
- Reduce the number of variables that will be used in the model: "**feature engineering**"



Step 4: Modeling

Select techniques to optimize performance

Exploratory Data Analytics & Data Visualization (50%)

Statistical and analytical methods that identify connections in the data, aiding human understanding and decision-making.

Unsupervised Machine Learning (20%)

Methods *learn* a structure that describes the relationships within a dataset. There are **no distinct labels** in the data; used for clustering or dimensionality reduction.

Supervised Machine Learning (20%)

Uses **labeled data** to develop a function that maps inputs to outputs. This *learned* function can then be applied to new data to predict an output variable.

Reinforcement Learning (10%)

Algorithm makes the optimal choice in a given situation, based on rewards. **Labels are discovered** by a process of trial and error.

Data Science in NASA Domains

Data Science

Step 5: Evaluation

Train, Validate and Test

Full Data Set

S	Var1	Var2	Var3	Label
0.001	10	5	0.817	0
0.002	9	1	0.993	1
0.003	10	3	0.972	0
0.004	4	2	0.630	0
0.005	10	2	0.007	0
0.006	6	5	0.231	0
0.007	10	3	0.143	0
0.008	7	4	0.995	0
0.009	5	4	0.319	0
0.01	5	4	0.139	0
0.011	2	5	0.612	0
0.012	6	2	0.955	1
0.013	5	2	0.561	1
0.014	5	2	0.949	1
0.015	9	1	0.070	0
0.016	10	4	0.932	0
0.017	7	2	0.637	1
0.018	10	3	0.803	0
0.019	6	5	0.944	0
0.02	8	5	0.725	0
0.021	4	5	0.859	0
0.022	6	4	0.195	0
0.023	2	5	0.147	0
0.024	10	4	0.159	0
0.025	3	1	0.984	0
0.026	9	4	0.856	0
0.027	9	1	0.749	1
0.028	2	4	0.350	0
0.029	8	1	0.544	1
0.03	5	3	0.724	0
0.031	9	2	0.374	0
0.032	10	2	0.220	0
0.033	2	5	0.262	0
0.034	8	3	0.378	0
0.035	9	3	0.084	0
0.036	7	5	0.141	0
0.037	5	3	0.090	0
0.038	4	2	0.103	0
0.039	8	1	0.011	0
0.04	3	3	0.413	0
0.041	8	1	0.823	1
0.042	7	5	0.034	0
0.043	6	1	0.644	1
0.044	7	5	0.121	0
0.045	2	1	0.698	0
0.046	10	4	0.199	0
0.047	2	3	0.891	0
0.048	7	1	0.355	0
0.049	8	1	0.757	1
0.05	4	3	0.534	0

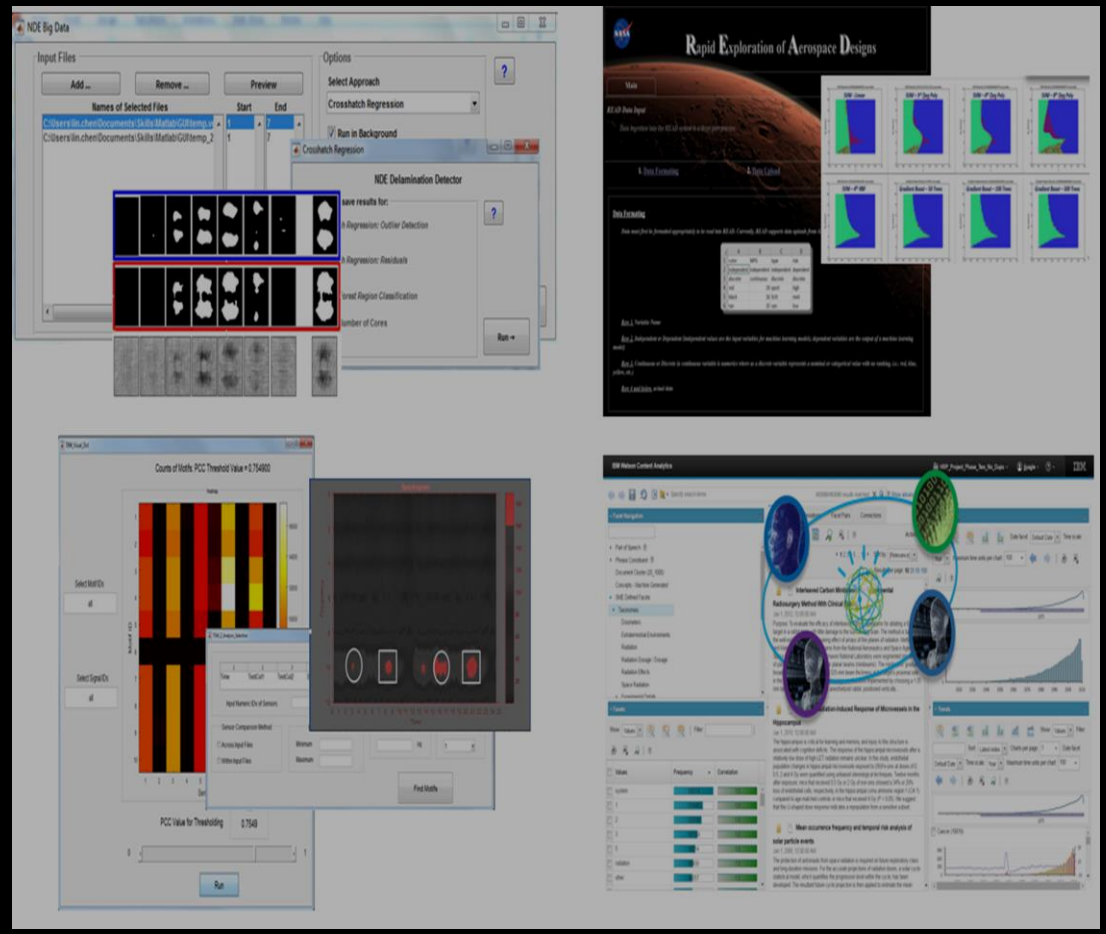
Training Set
Used to develop potential functions during the learning phase

Validation Set
Used to evaluate the performance of developed functions, and calibrate for improvement

Test Set
Used with the final model to report accuracy and assess generalization

Step 6: Deployment

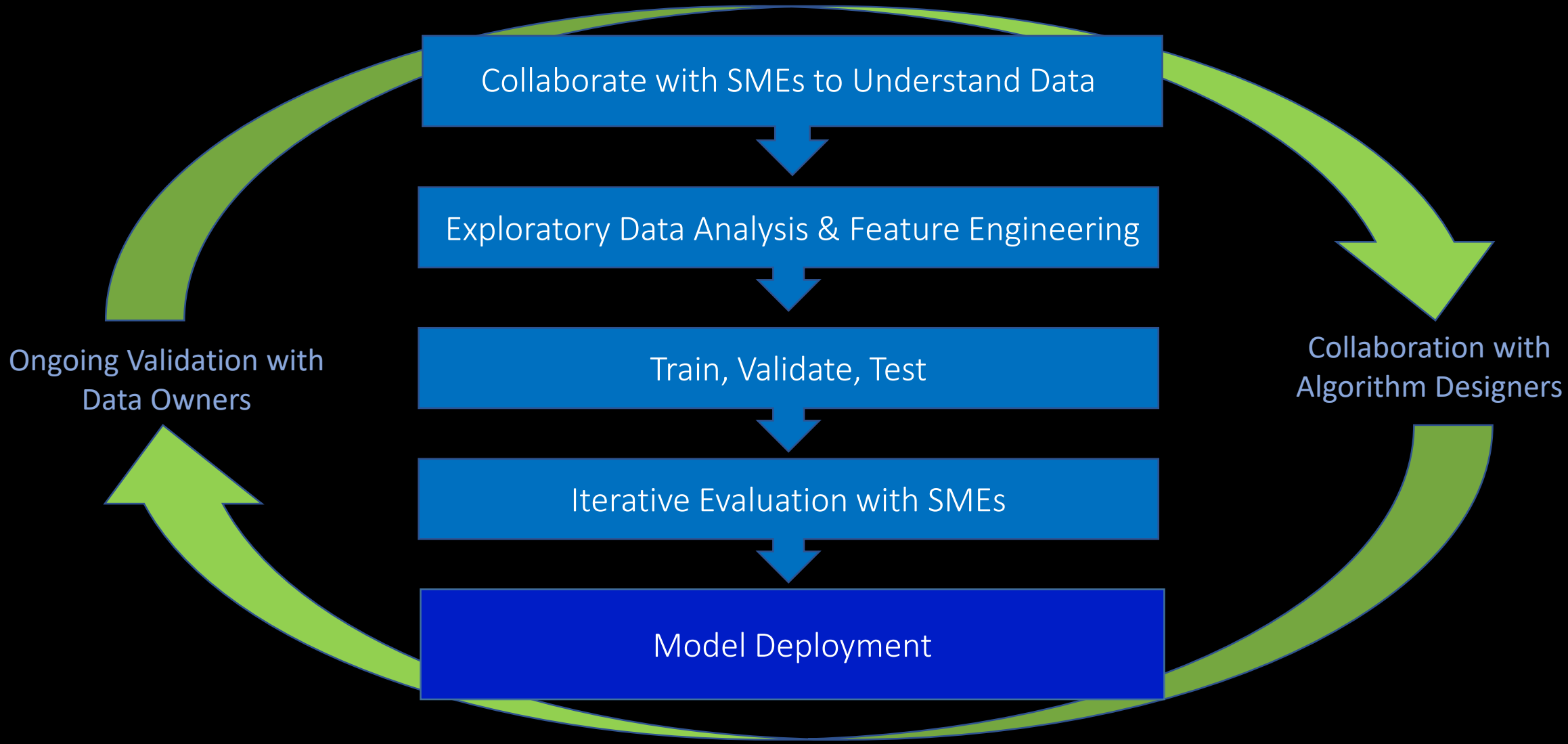
Design user interface or configure existing tools



Data Science Workflow

Validate Approach and Methods with Subject Domain Experts (SMEs)

Data Science



Data Science Use Cases

Data Science

Knowledge Discovery in Databases

- Efficient analysis of mod-sim data
- Detecting deviations in travel patterns

Data Mining

- Detecting patterns in trajectory data from STS missions
- Understanding the origin of defects in additively manufactured components



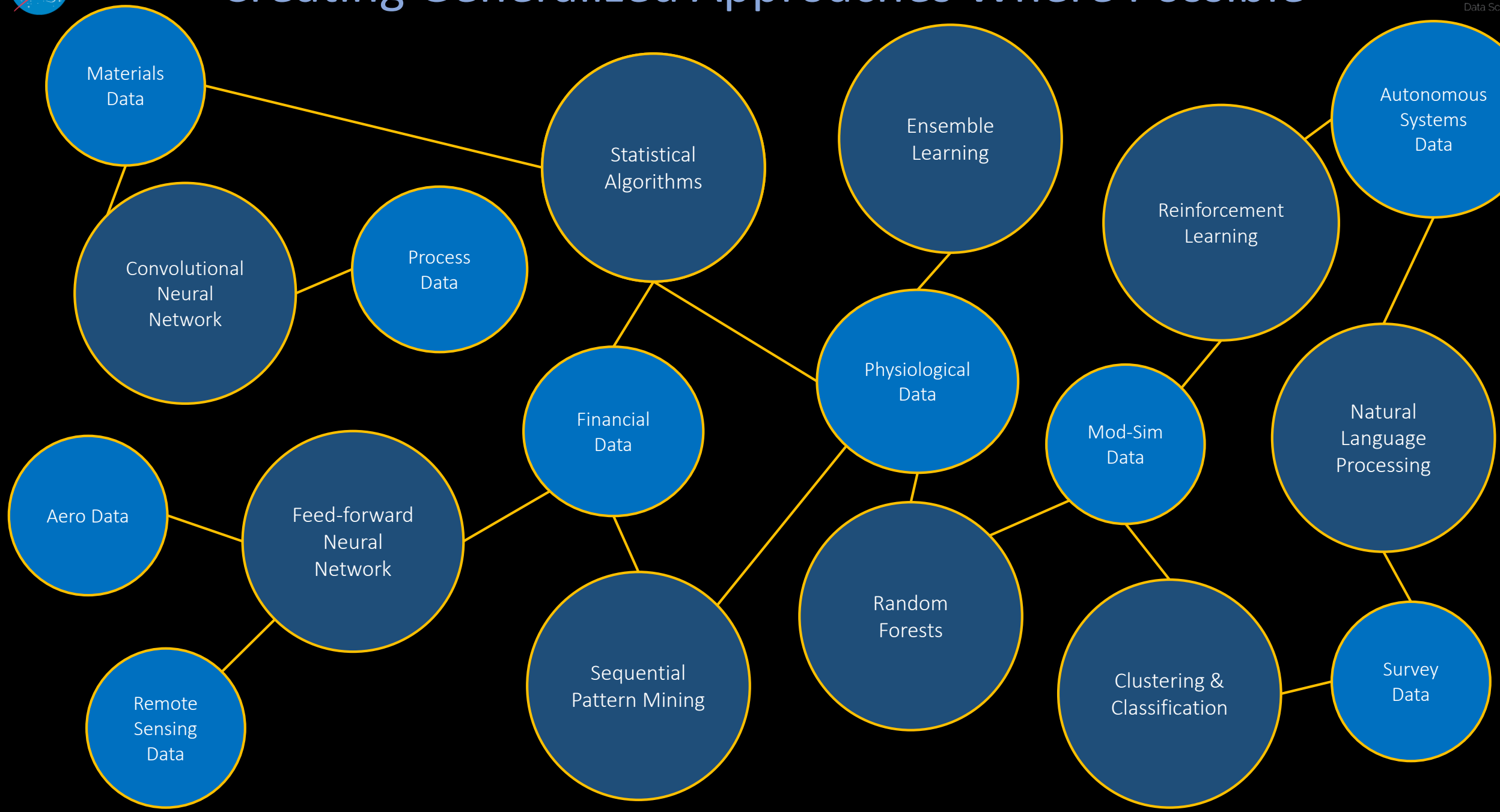
Machine Learning

- Detecting delaminations and cracking in carbon fiber
- Predicting pilot cognitive state during flight
- Defining areas of operability for aircraft
- Predicting severe weather events
- Creating multi-fidelity models to use compute resources more intelligently

Reinforcement Learning

- Exploring reward functions
- Identifying reachable states

Creating Generalized Approaches Where Possible



Capability Through Collaboration



Data Science

A balanced approach to teaming, consultation, and education, to grow an embedded digitally-enabled workforce at Langley

DO

TEAM

CONSULT