Geospatial Data Science:
An Experiential Journey

Manil Maskey, PhD
IEEE GRSS Earth Science Informatics Technical Committee Co-chair
NASA Science Mission Directorate
NASA IMPACT/MSFC

Team

Brian Freitag
NASA

Iksha Gurung
UAH

Muthukumaran
Ramasubramanian
UAH

Sean Harkins
Development Seed

Johannes Jakubik
IBM Research

Linsong Chu
IBM Research

Paolo Fraccaro
IBM Research

Blair Edwards
IBM Research

Manil Maskey
NASA

Goals

Continue the summer school series: building capacity around data science

Provide full scope of data science: From Data Generation to Processing, Analysis, and Applications

Explore research and applications of Geospatial Foundation Models

Provide hands on experience:

- Use cloud native tools for interactive analysis
- Fine-tune geospatial foundational models for practical applications
- Use fine-tuned models for inferences and visualizations

Provide forum to exchange Ideas

Foster collaboration

Data Science Overview

Definition(s)

Narrow: extracting knowledge and insights from data

"Today, the term 'data scientist' typically describes a knowledge worker who uses the complex and massive data resources characteristic of this new era. However, *data science is a broader concept involving principles for data collection, storage, integration, analysis, inference, communication, and ethics appropriate for this new data-driven era.*"

National Academies of Sciences, Engineering, and Medicine 2018. Data Science for Undergraduates: Opportunities and Options. Washington, DC: The National Academies Press. https://doi.org/10.17226/25104.
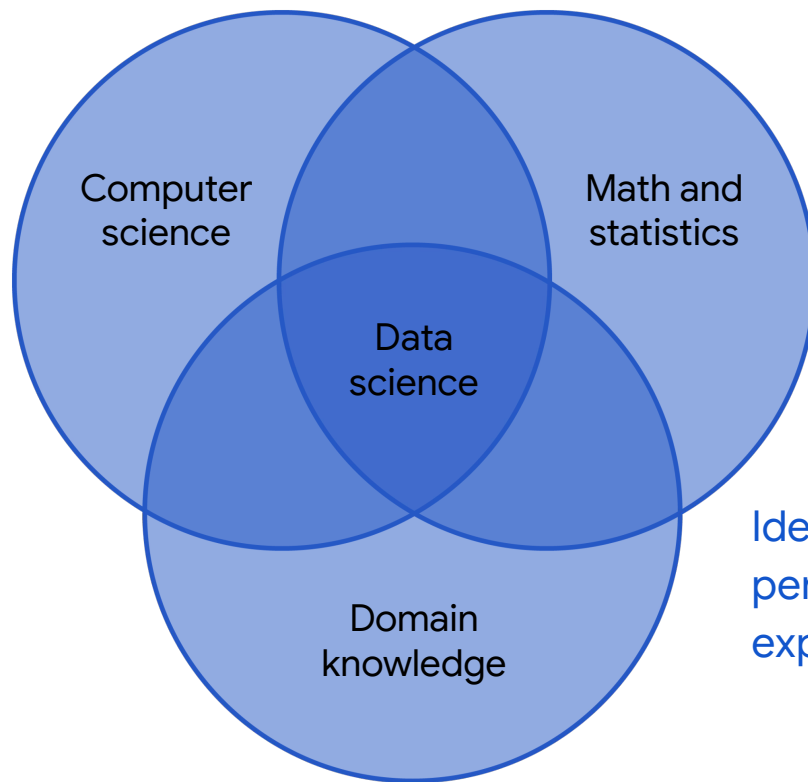
Insight

Information
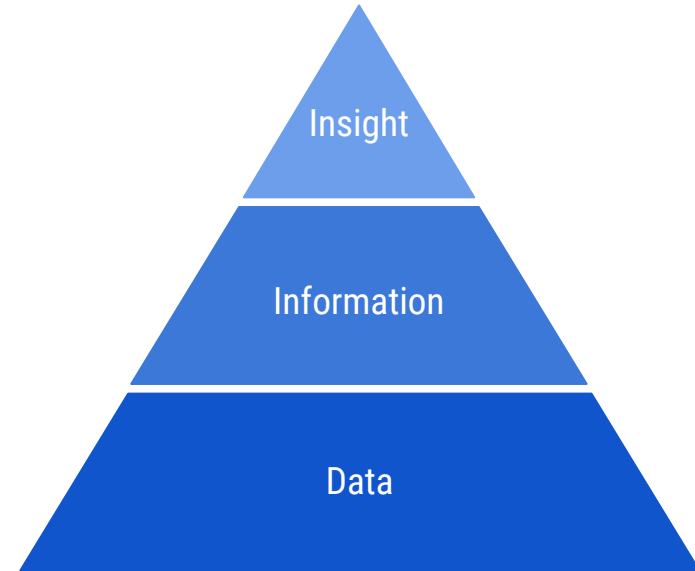
Data

Complex Data and Research Life Cycle

Data science is about (data intensive) process, and the process can be complex depending on the specific science domain

Data Science: Unique Skillset

Computer science

Math and statistics

Data science

Domain knowledge

Ideal skills needed by a person to be a data science expert.

Skills Required

Programming computers

Working with data

Descriptive statistics

Data visualization

Statistical modeling

Handling big data

Machine learning

Deploying to production
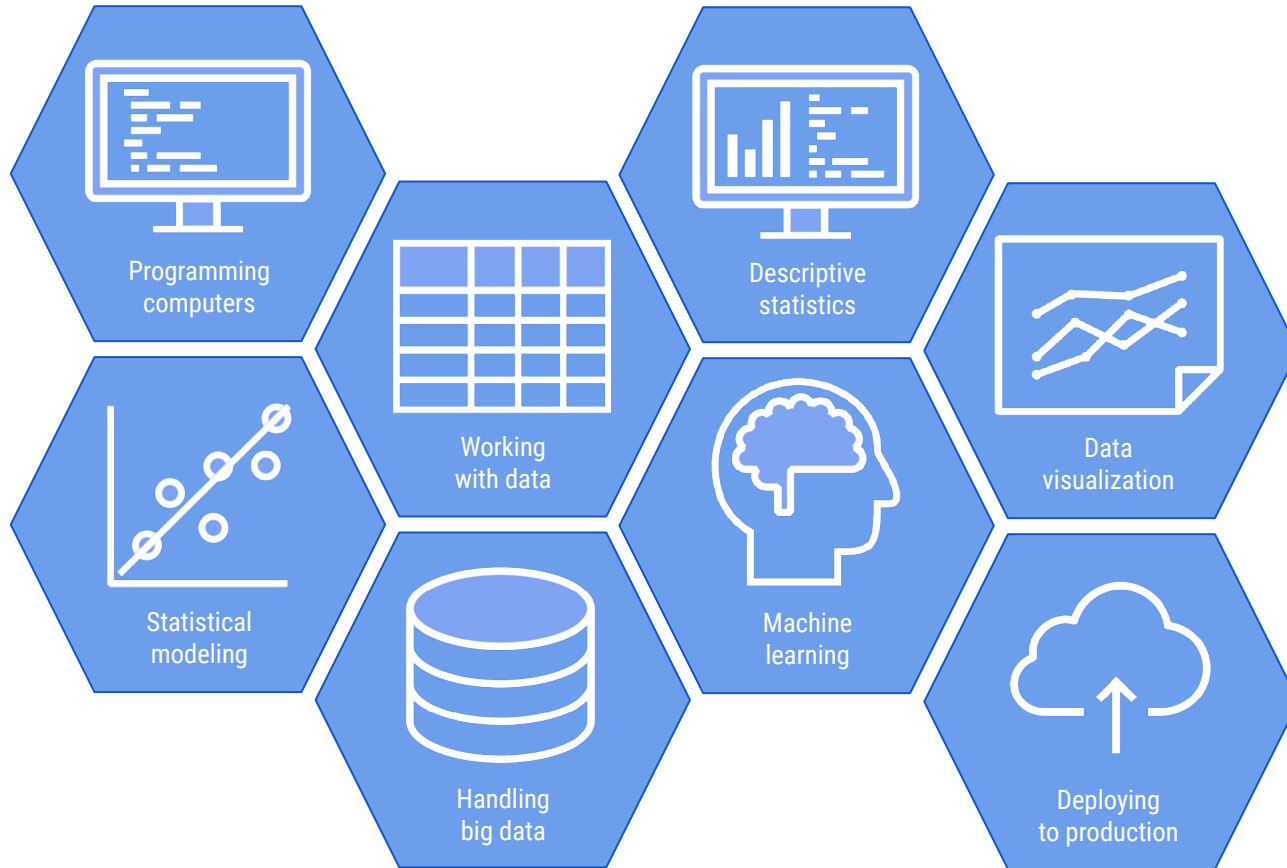
Interconnected teams are the answer.

Data Science: Geospatial Domain

NASA EARTH FLEET
OPERATING & FUTURE THROUGH 2023

SWOT (CNES)
LANDSAT-9 (USGS)   SENTINEL-6 Michael Freilich/B (ESA)
TROPICS (6)   GEOCARB
NISAR (ISRO)   MAIA
TSIS-2   TEMPO
PREFIRE (2)   PACE (NSO)
GLIMR   ICESAT-2
                                      GRACE-FO (2) (DLR)
                                      CYGNSS (8)
                                      NISTAR, EPIC (DSCOVR/NOAA)
                                      CLOUDSAT (CSA)
ISS INSTRUMENTS   TERRA (JAXA, CSA)
                                      AQUA (JAXA, AEB)
EMIT   AURA (NSO, FMI, UKSA)
CLARREO-PF   CALIPSO (CNES)
GEDI   GPM (JAXA)
SAGE III   LANDSAT 7 (USGS)
OCO-3   LANDSAT 8 (USGS)
TSIS-1
ECOSTRESS   OCO-2
LIS   SMAP
                                      SUOMI NPP (NOAA) (JAXA)
JPSS-2, 3 & 4 INSTRUMENTS

OMPS-Limb
LIBERA

03.24.20

INVEST/CUBESATS

RainCube
CSIM-FD
CubeRRT
TEMPEST-D
CIRiS
HARP
CTIM
HyTI
SNoOPI
NACHOS

(PRE) FORMULATION ●
IMPLEMENTATON ●
PRIMARY OPS ●
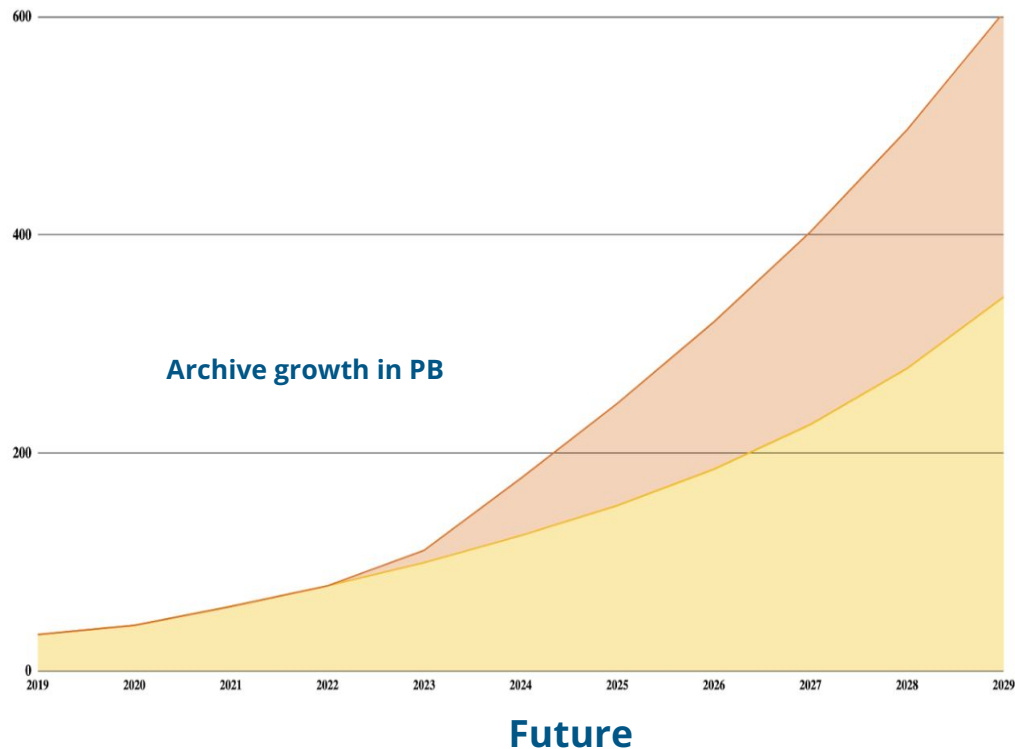EXTENDED OPS ●

10

By the number

62 PB Archive	**2 Billion** Data products
3.56 Billion Files in archive	**1.7 Million** Distinct data users

Now

Archive growth in PB

Future

Science of "Data Science"

Policy

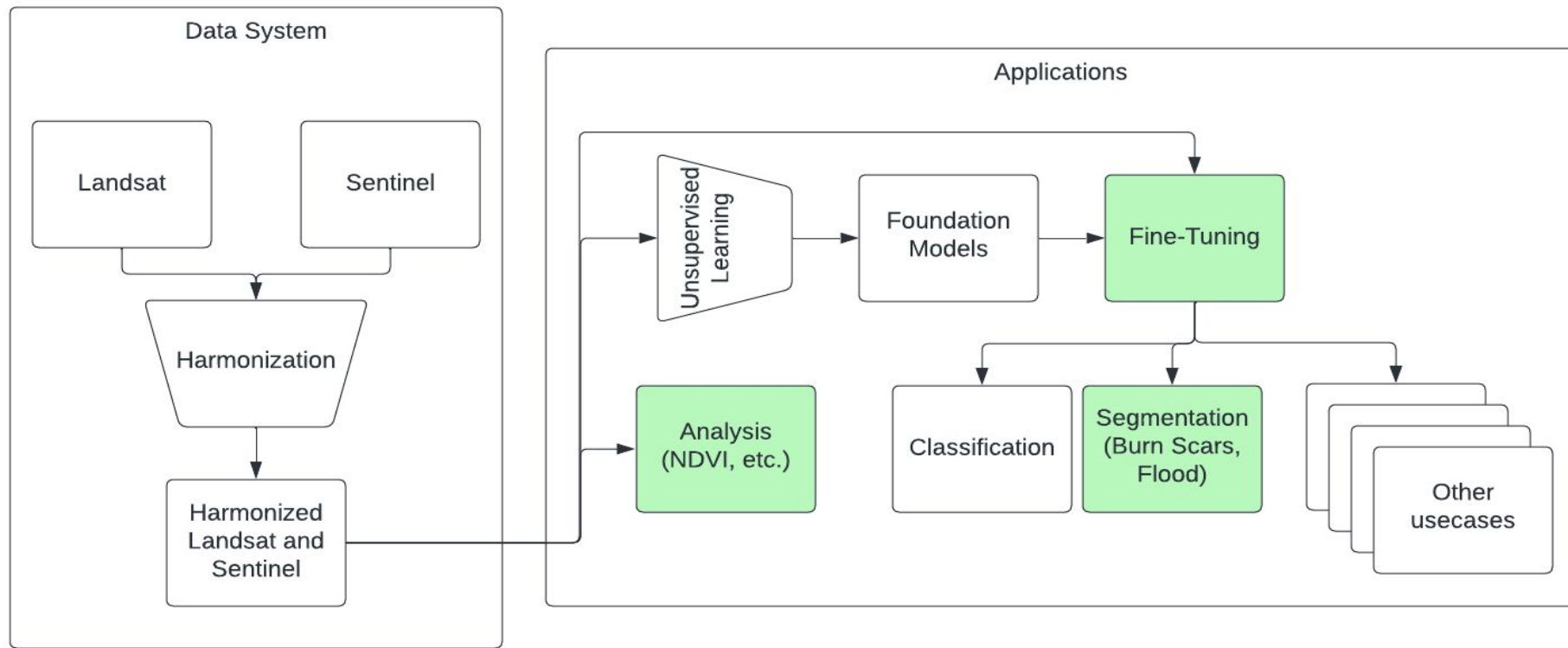Infrastructure

Research & Application Partnership

Community

What we hope to do today

- Chapter 1: Data Production/processing
    - Large Scale Data Harmonization
- Chapter 2: Data Analysis
    - Tools NASA FIRMS (HLS applications and dynamic tiling capabilities)
    - Interactive HLS notebook for analysis and visualization
- Chapter 3: Theory & Application of Geospatial Foundation Model
    - Fine-tune HLS foundation model for specific use-cases: Flood and Burn Scars
- Chapter 4: Interactive Exploration of Fine-tuned Model

Data systems and applications

Building Blocks: Future State
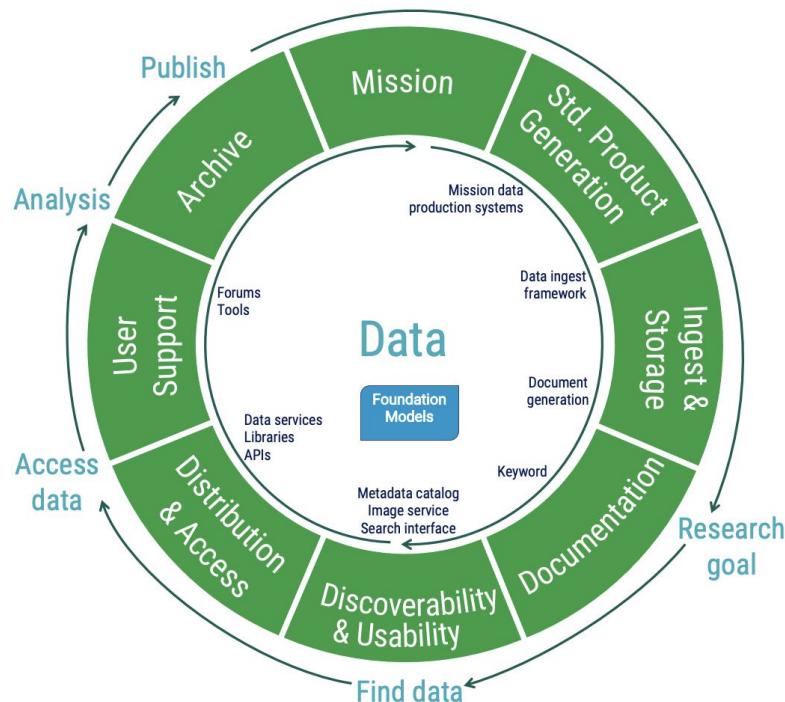
Data + Tools/Infrastructure
+
AI Foundational Models
>>>
Accelerate Research and Applications

Takeaway: Expanding Proficiency in Geospatial Data Science

1. Data science extends beyond applying ML/AI algorithms; it encompasses a scientific field with its own processes.

2. The data science process can be intricate, varying based on the particular scientific domain.

3. Gain comprehension of the geospatial data science lifecycle.

4. Acquire knowledge in handling large scale data production.

5. Establish a data science environment and conduct interactive analysis.

6. Develop a strong understanding in geospatial foundation models.

7. Apply a finely-tuned geospatial foundation model effectively.

Thank you

manil.maskey@nasa.gov

- IEEE GRSS
- Earth Science Informatics TC
- HDCRS WG
- University of Iceland
- IBM Research
- Development Seed
- All the participants