



Geospatial Data Science: An Experiential Journey

Manil Maskey, PhD
NASA Science Mission Directorate
NASA IMPACT/MSFC

IEEE GRSS HDCRS Summer School
IEEE GRSS High Performance and Disruptive Computing in Remote Sensing Summer school
29 May - 1 June, 2023, Reykjavik, Iceland



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

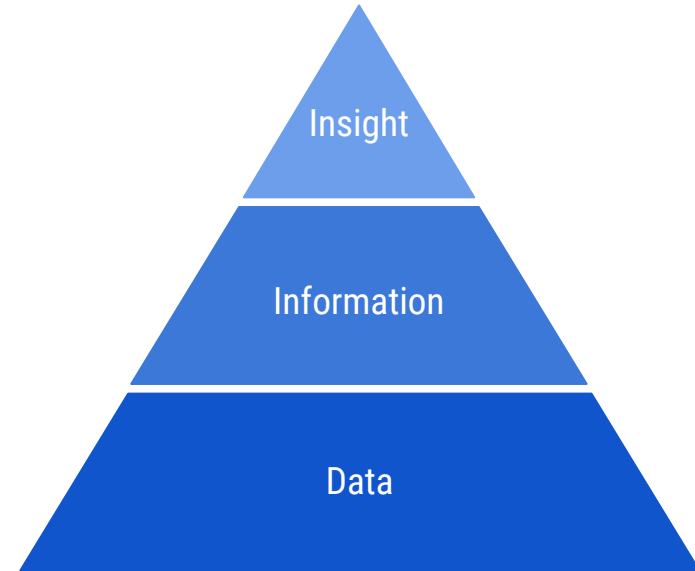
Data Science Overview

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>.*

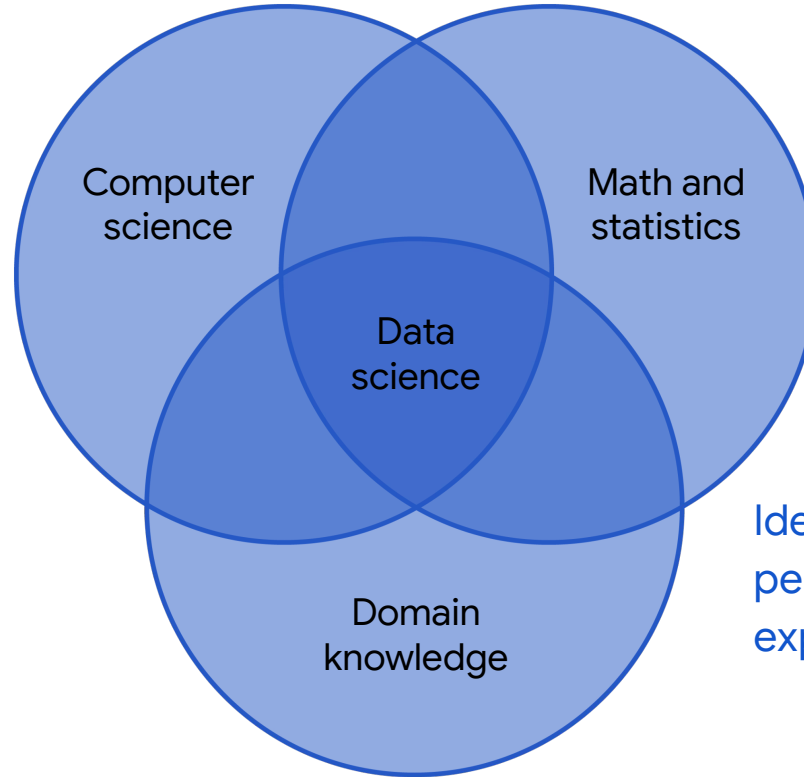


Complex Data and Research Life Cycle

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

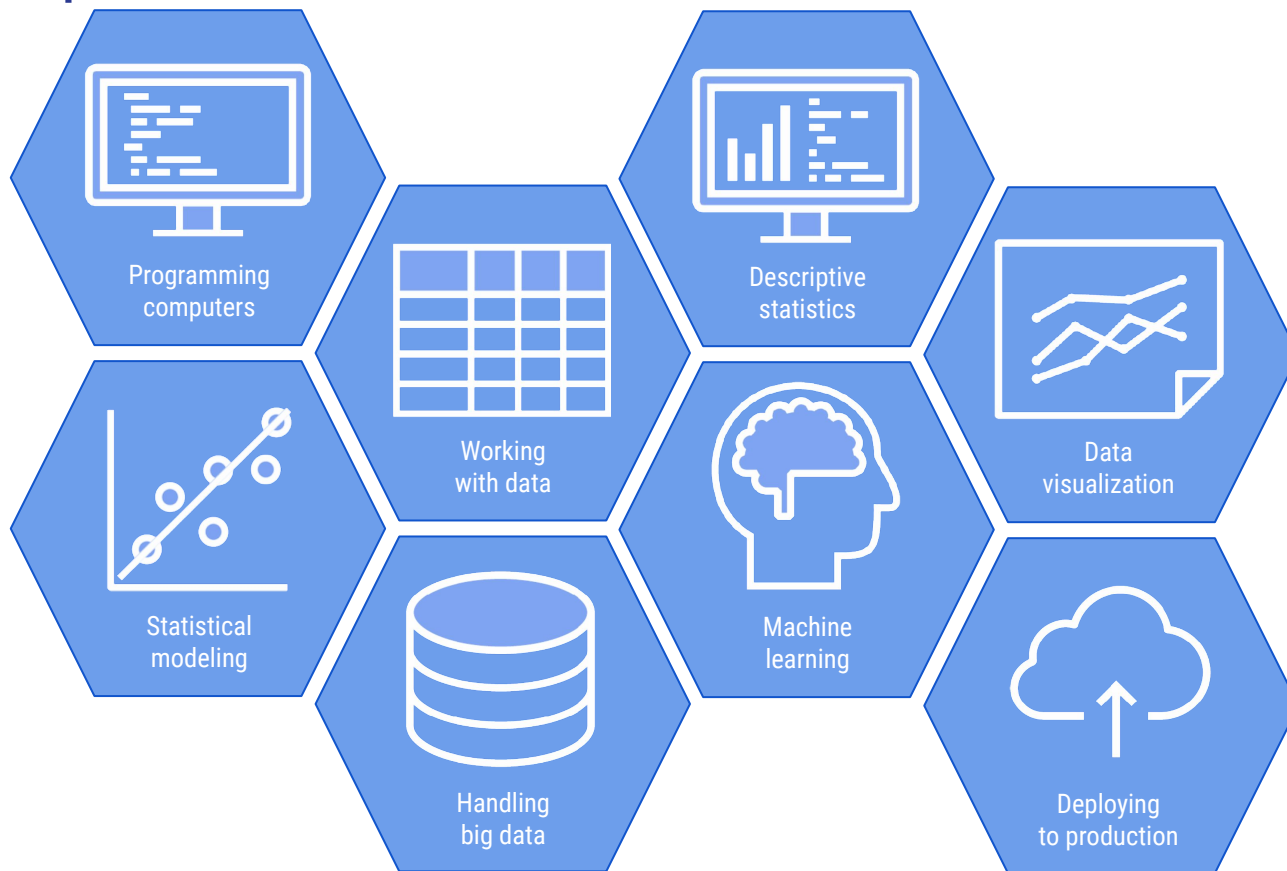


Data Science Skills



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

Skills Required



Interconnected
teams are the
answer.

Data Science: Geospatial Domain

NASA EARTH FLEET

OPERATING & FUTURE THROUGH 2023

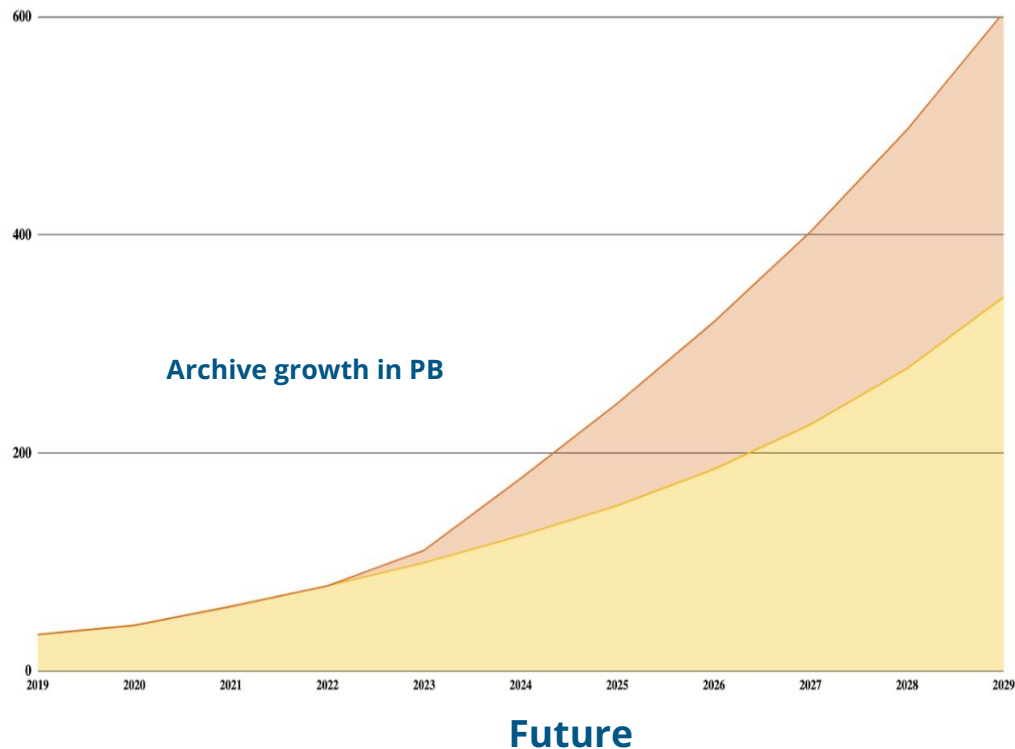
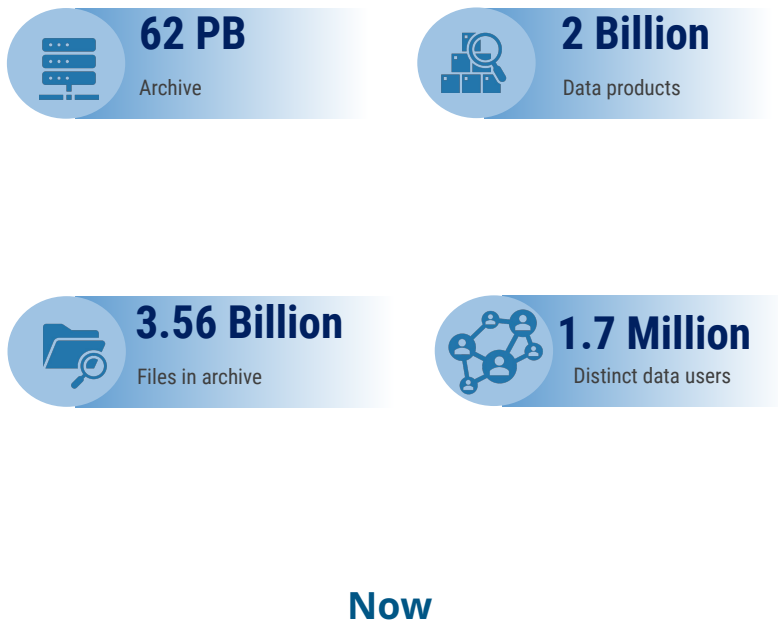


INVEST/CUBESATS

RainCube
CSIM-FD
CubeRRR
TEMPEST-D
CIRIS
HARP
CTIM
HyTI
SNoOPI
NACHOS

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

By the number



Science of “Data Science”

Policy

Infrastructure

Research &
Application
Partnership

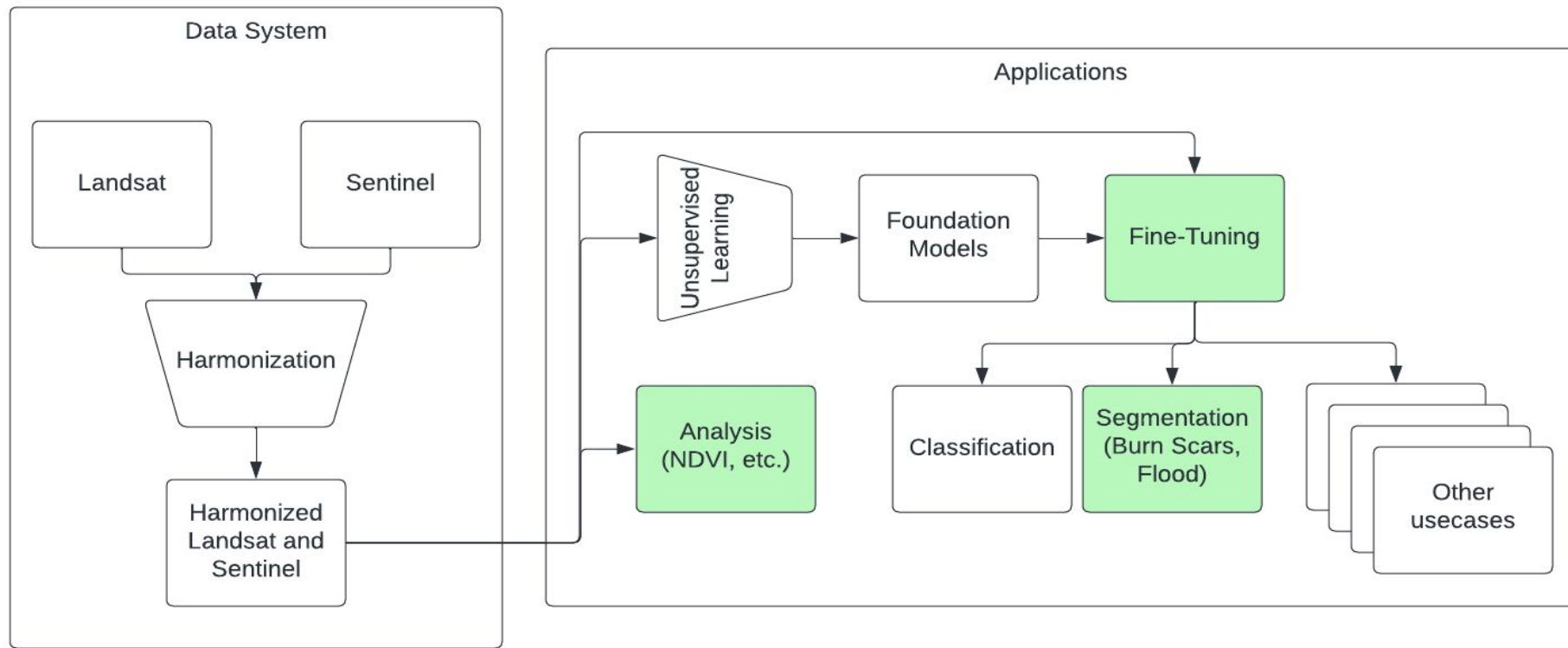
Community



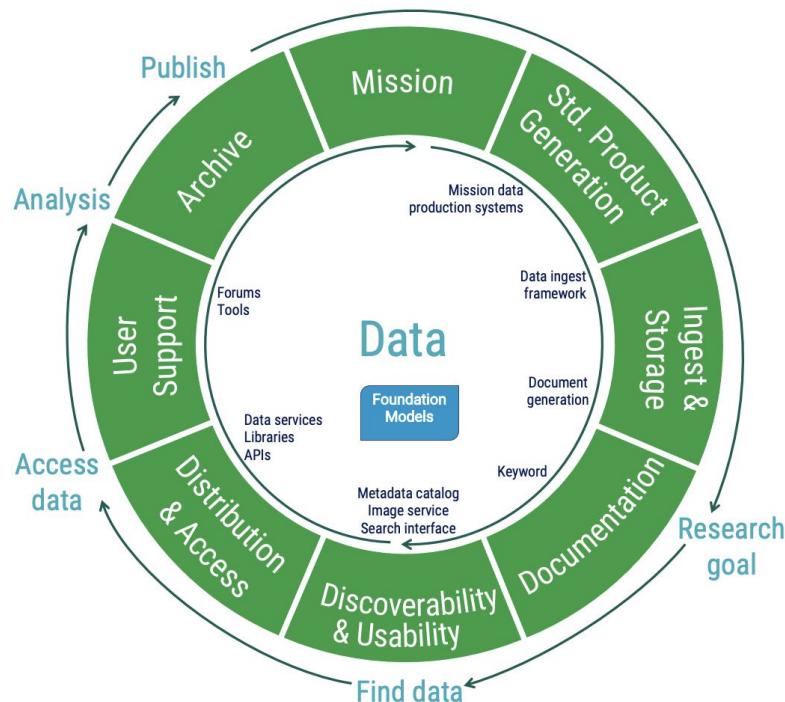
What we hope to do today

- 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
- University of Iceland
- Earth Science Informatics TC
- IBM Research
- Development Seed
- All the participants