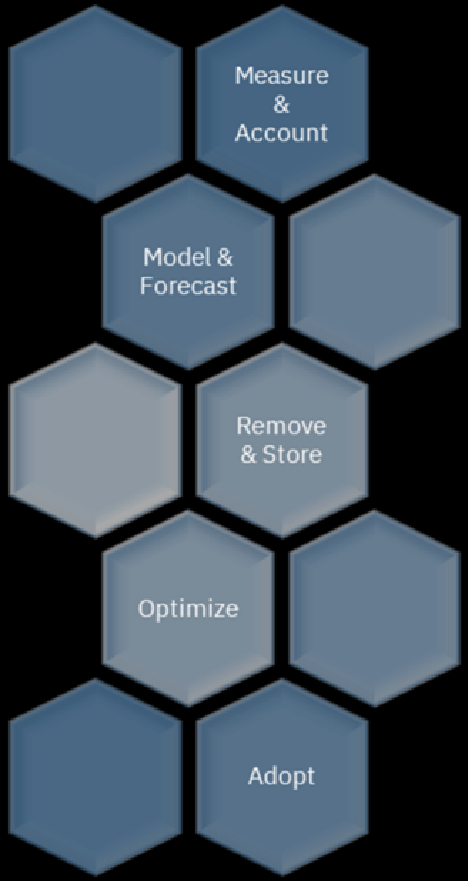
Harnessing the Geospatial Science Revolution
to Advance Sustainability Research:

A Geospatial Discovery Network

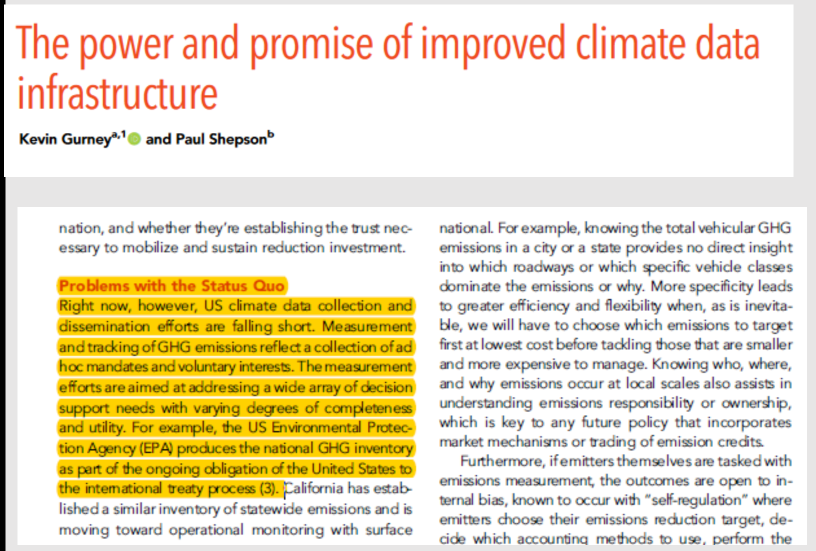
C.D. Watson, H.F. Hamann, K. Weldemariam,
B. Edwards, A. Jones, and J. Schmude

Accelerating computing for climate is critical for battling climate change



- *Greenhouse Gas Monitoring and Analysis*
- *Accelerating ESG reports*
- *Surrogate climate models*
- *AI for Materials Discovery*
  - *Carbon Capture*
  - *Membranes*
  - *Battery Materials*
- *Natural Carbon Sequestration*
- *Smarter transportation, buildings*
- *Next gen electric grid*
- *Climate impact modeling*
  - *Flood, wildfire, draughts, heatwaves etc.*
  - *Socio-economic*
- *Risk and impact discovery*
- *…*

Traditional systems are outpaced by explosion of data, complexity and urgency of the matter



The power and promise of improved climate data infrastructure

Kevin Gurney[a,1] and Paul Shepson[b]

nation, and whether they're establishing the trust necessary to mobilize and sustain reduction investment.

Problems with the Status Quo

Right now, however, US climate data collection and dissemination efforts are falling short. Measurement and tracking of GHG emissions reflect a collection of ad hoc mandates and voluntary interests. The measurement efforts are aimed at addressing a wide array of decision support needs with varying degrees of completeness and utility. For example, the US Environmental Protection Agency (EPA) produces the national GHG inventory as part of the ongoing obligation of the United States to the international treaty process (3). California has established a similar inventory of statewide emissions and is moving toward operational monitoring with surface

national. For example, knowing the total vehicular GHG emissions in a city or a state provides no direct insight into which roadways or which specific vehicle classes dominate the emissions or why. More specificity leads to greater efficiency and flexibility when, as is inevitable, we will have to choose which emissions to target first at lowest cost before tackling those that are smaller and more expensive to manage. Knowing who, where, and why emissions occur at local scales also assists in understanding emissions responsibility or ownership, which is key to any future policy that incorporates market mechanisms or trading of emission credits.

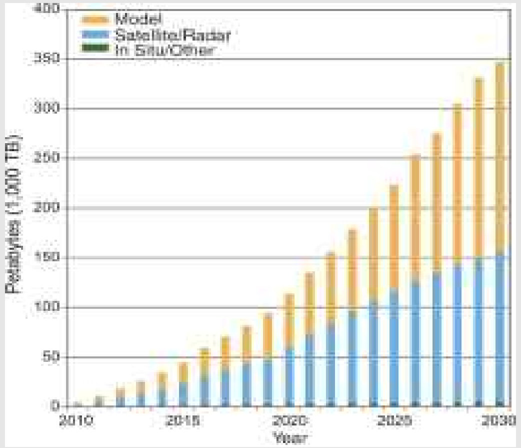
Furthermore, if emitters themselves are tasked with emissions measurement, the outcomes are open to internal bias, known to occur with "self-regulation" where emitters choose their emissions reduction target, decide which accounting methods to use, perform the

Opinion: The power and promise of improved climate data infrastructure by Kevin Gurney, Paul Shepson *Proceedings of the National Academy of Sciences* 118.35 (2021).

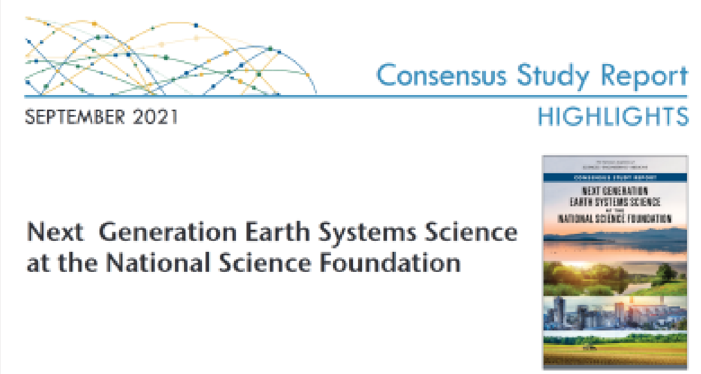
Climate Data Challenges in the 21st Century

Abstract

Climate data are dramatically increasing in volume and complexity, just as the users of these data in the scientific community and the public are rapidly increasing in number. A new paradigm of more open, user-friendly data access is needed to ensure that society can reduce vulnerability to climate variability and change, while at the same time exploiting opportunities that will occur.



Climate Data Challenges in the 21st Century
By Jonathan T. Overpeck, Gerald A. Meehl, Sandrine Bony, David R. Easterling *Science* 11 700 (2001)



Use observational, computational, and modeling capabilities synergistically to accelerate discovery and convergence. The observational, computational, and modeling infrastructure must work collectively to support convergence in Earth Systems Science. Observations and monitoring reveal changes in the Earth's systems. Data from diverse sources are assimilated into models that rep-resent natural- and social-system processes and their interactions across the Earth's systems. Computation provides the framework for putting together the complex pieces of Earth Systems Science, supporting data collection and analysis, generation of forecasts, and interpretation of model results.

Key challenges are routed in the vast logical and physical distribution of climate-relevant information

USGS science for a changing world

THE WORLD BANK IBRD • IDA

NASA SEDAC

IBM

NASA EARTHDATA OPEN ACCESS FOR OPEN SCIENCE

Microsoft Planetary Computer

African Development Bank / African Development Fund

ECMWF

The Amazon Sustainability Data Initiative ASDI

United States Census Bureau

Google Earth Engine

NOAA

UNEP

Copernicus Europe's eyes on Earth

International Monetary Fund

DATA.GOV

CLIMATE CHANGE RESOURCE CENTER

ESGF Earth System Grid Federation

GeoNetwork opensource

Environment and Climate Change Canada

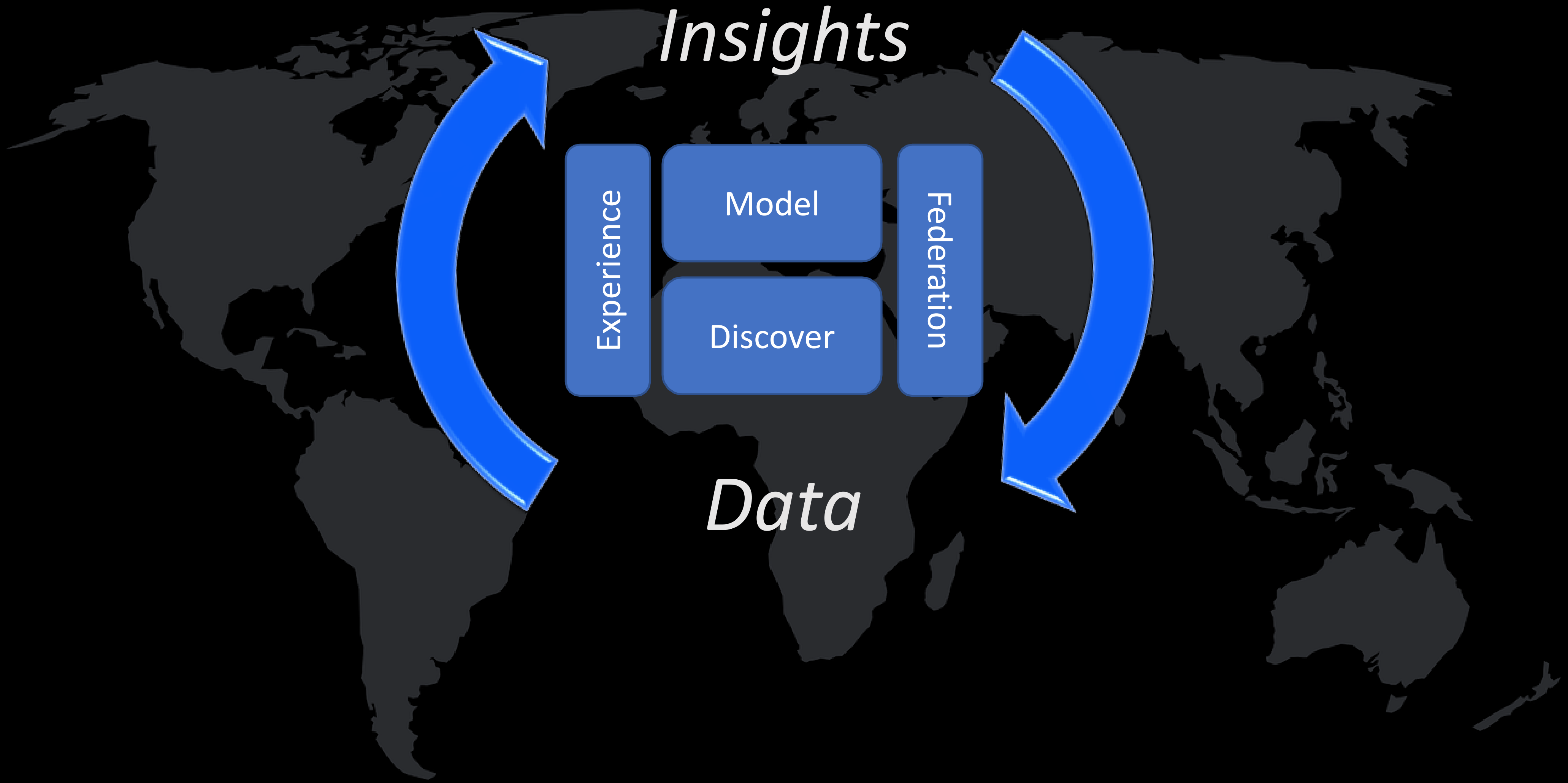
Data gravity, lack of discoverability and modeling tools are amplifying the challenge

Modeling frameworks

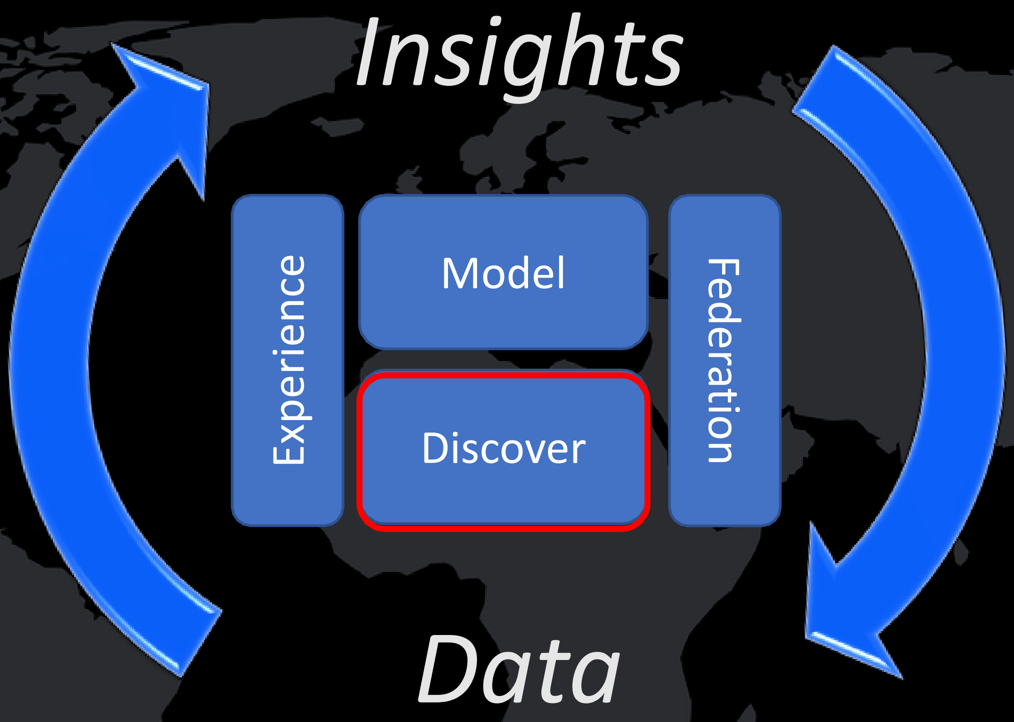
Discoverability

Data Gravity

A Geospatial Discovery Network for distributed and federated discovery and modeling



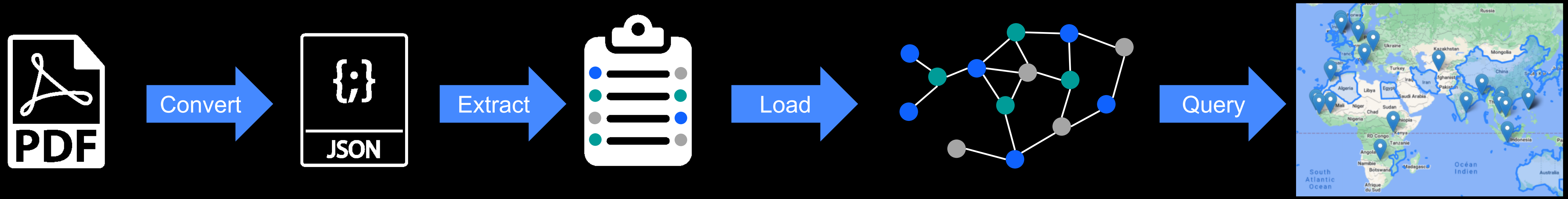
A Geospatial Discovery Network for distributed and federated discovery and modeling



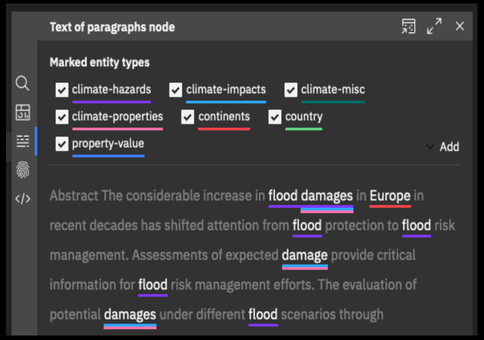
1. Dataset Discovery (state of the art))
2. Knowledge Discovery
3. Pixel-level Discovery
4. Feature Discovery

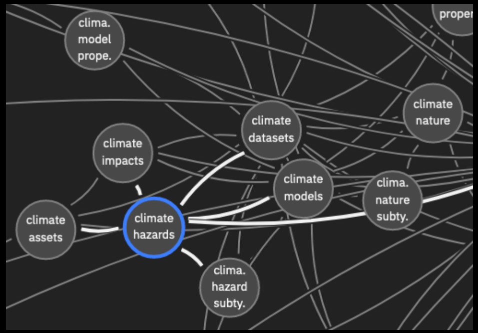
Accelerated Geospatial Knowledge Discovery with Deep Search

Document Conversion into Machine Readable Output	Climate Entity & Relationship Recognition	Climate Knowledge Representation	Geo-Spatial Document Visualization



Knowledge Corpus
- 1 Mio. Climate related Abstracts
- IPCC Reports
- 8k ESG Reports
- 2 Mio. arXiv Publications



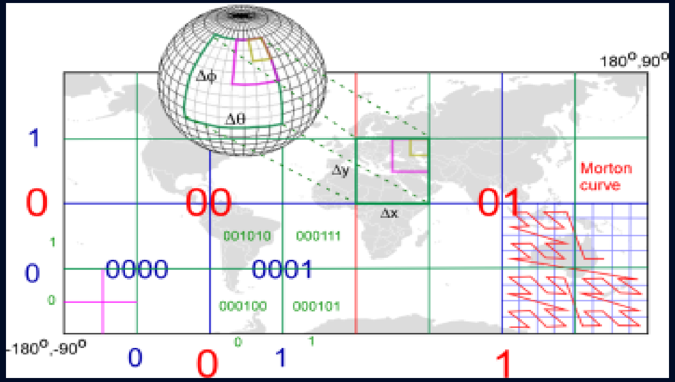


Geospatial Knowledge Discovery

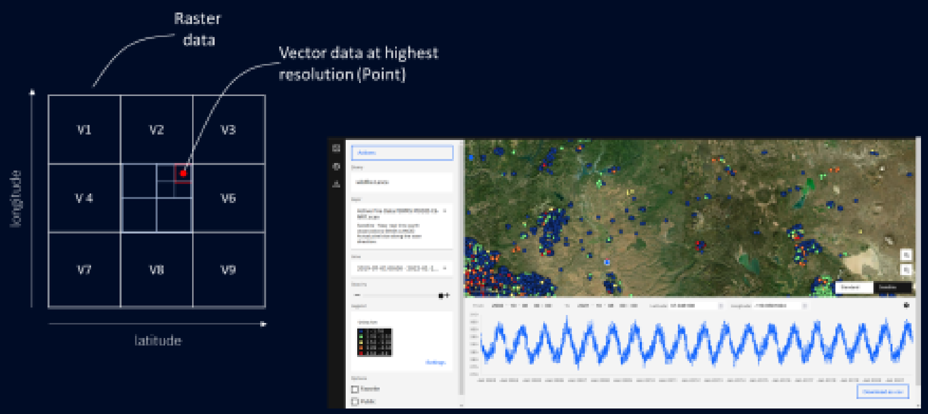
For a location and time:
- which surveys?
- what impact functions?
- what observations?
- what flood models?

Accelerated Pixel-level Discovery

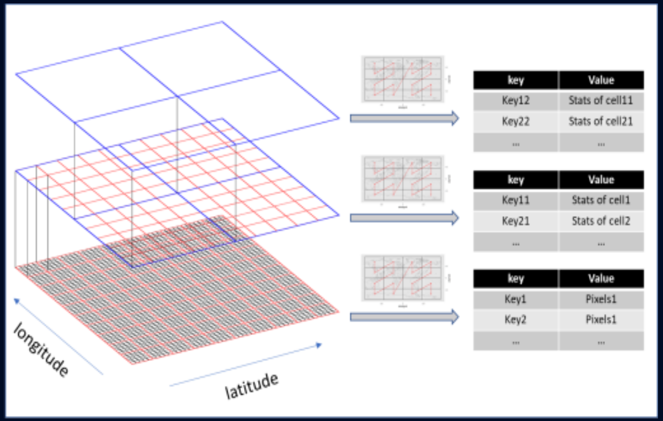
1. Nested resolution layers & common coordinate systems



3. Raster-Vector Data Cubes

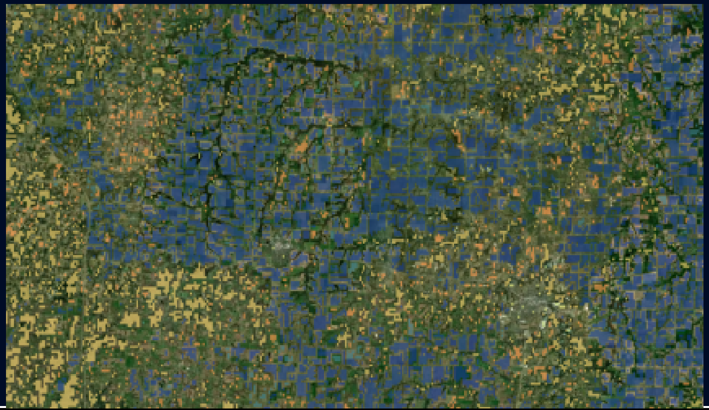


2. Data/Pixel-level indexing with overview layers



⇒ Vector-Raster Data-level Discovery

(e.g. find areas and timestamps with high population and precipitation rates 5x than climatology?)

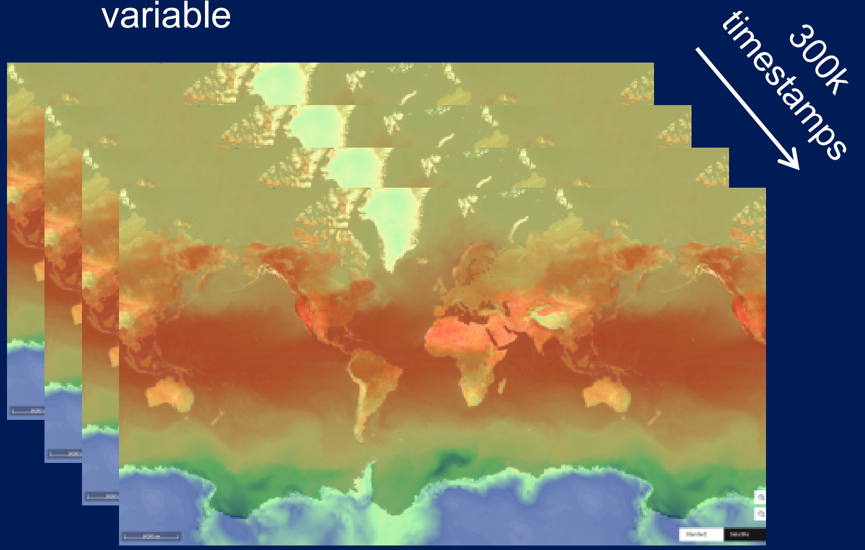


Accelerated Feature Discovery

Example: Spatially-resolved surface temperature gradients

1. Available data

4.5M locations* with 300,000 timestamps
4.5M * 4 Bytes * 300,000 ~ 5.4 Terabytes per variable



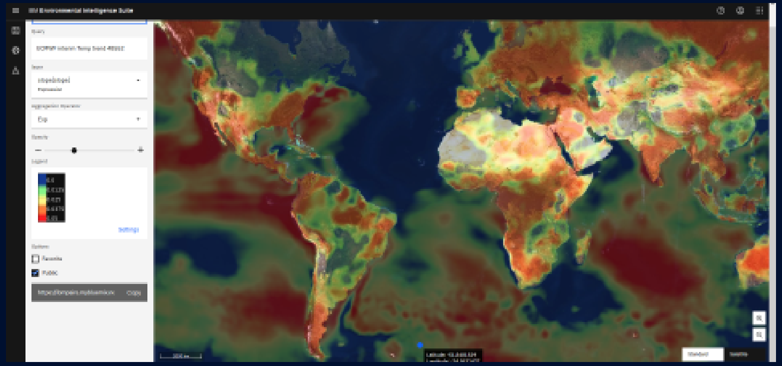
2. User-defined functions

supporting arbitrary math to be submitted without downloads

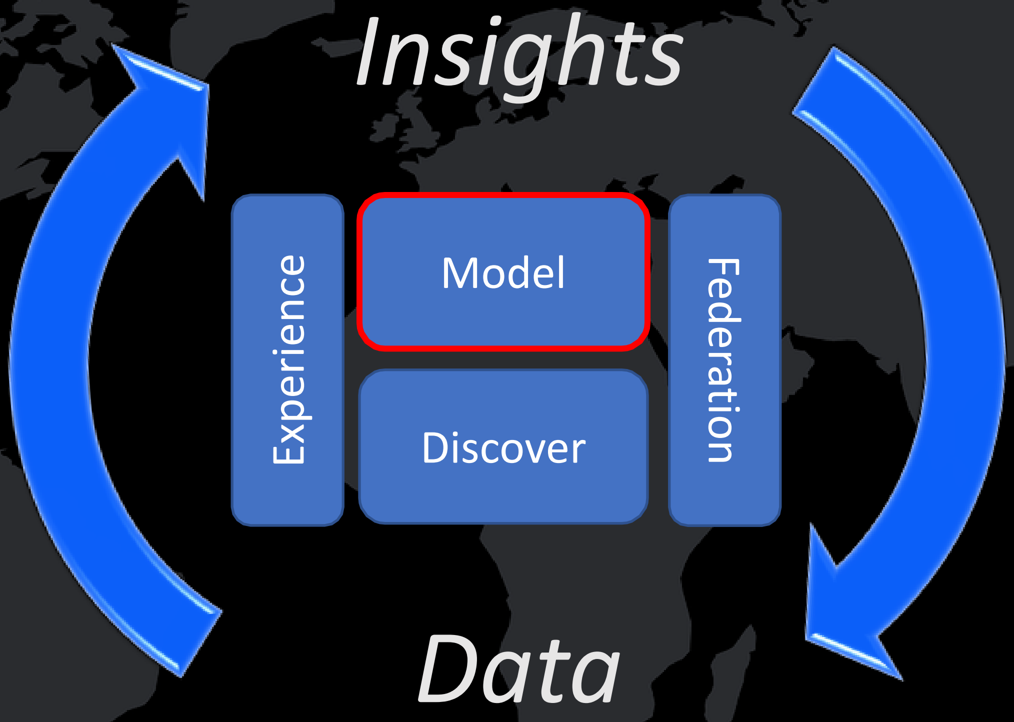
Time-series analysis (regression) over 300k timestamps for 4.5M locations

3. Features created

in 20 mins on a single thread

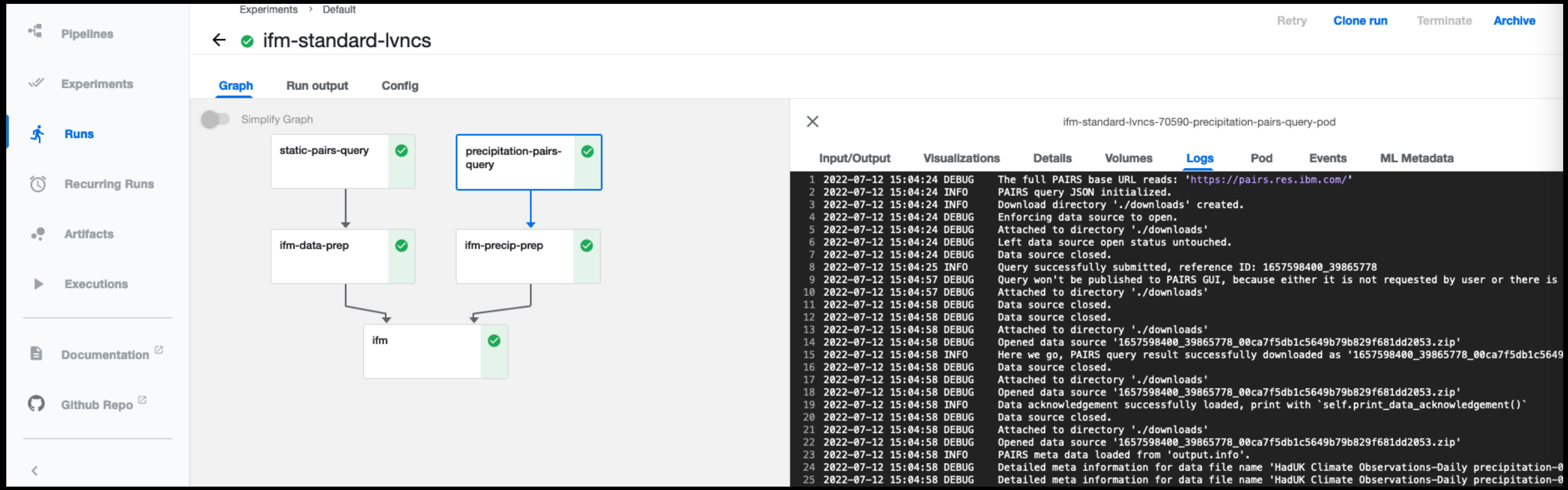


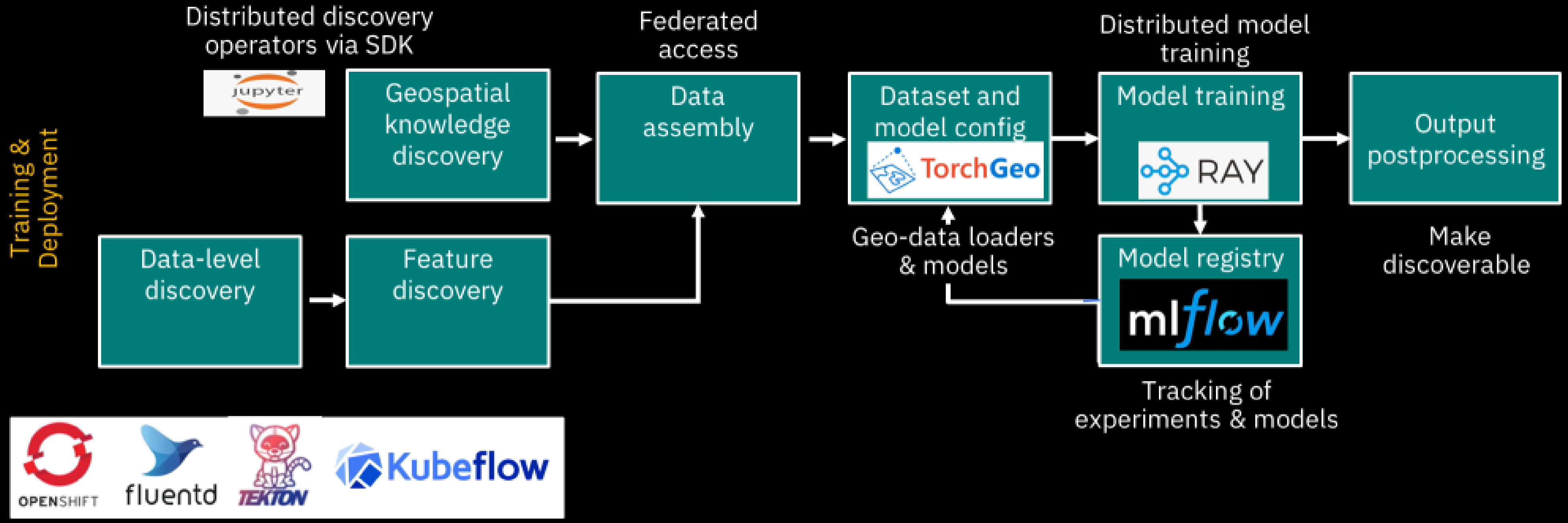
A Geospatial Discovery Network for distributed and federated discovery and modeling



1. Model modularization
2. Foundation Model

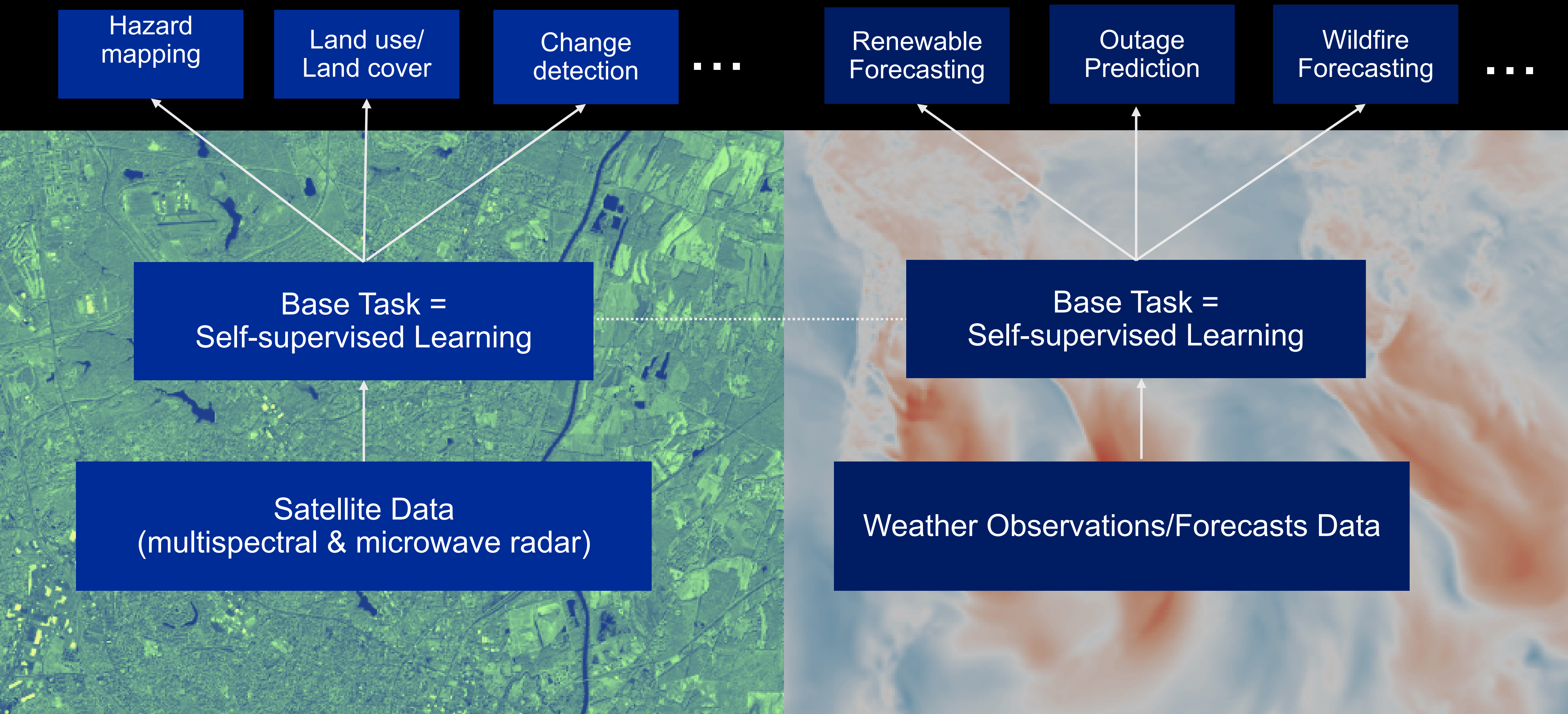
Modular Modeling Framework



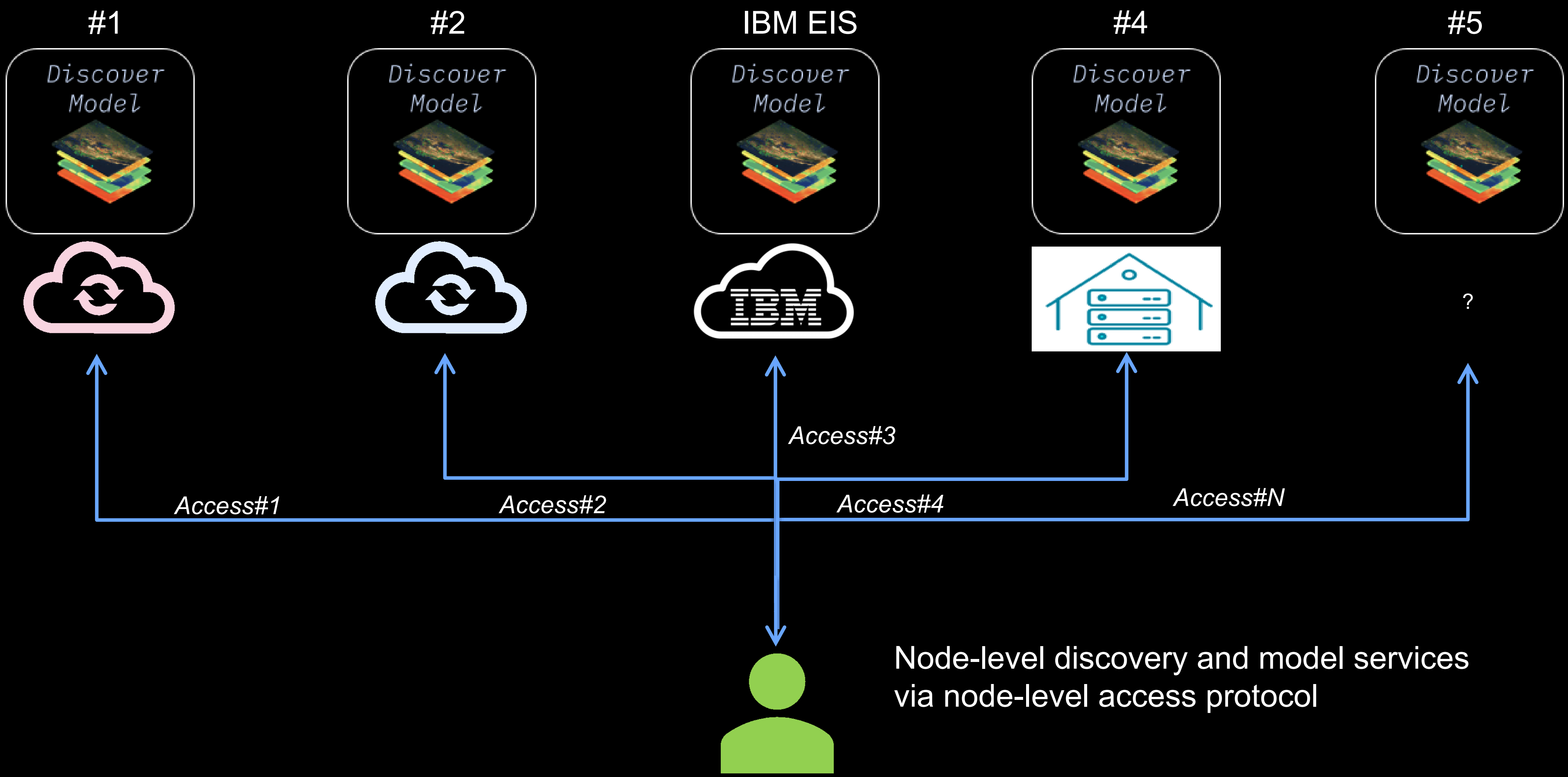


Geospatial Foundation Models

Downstream Tasks

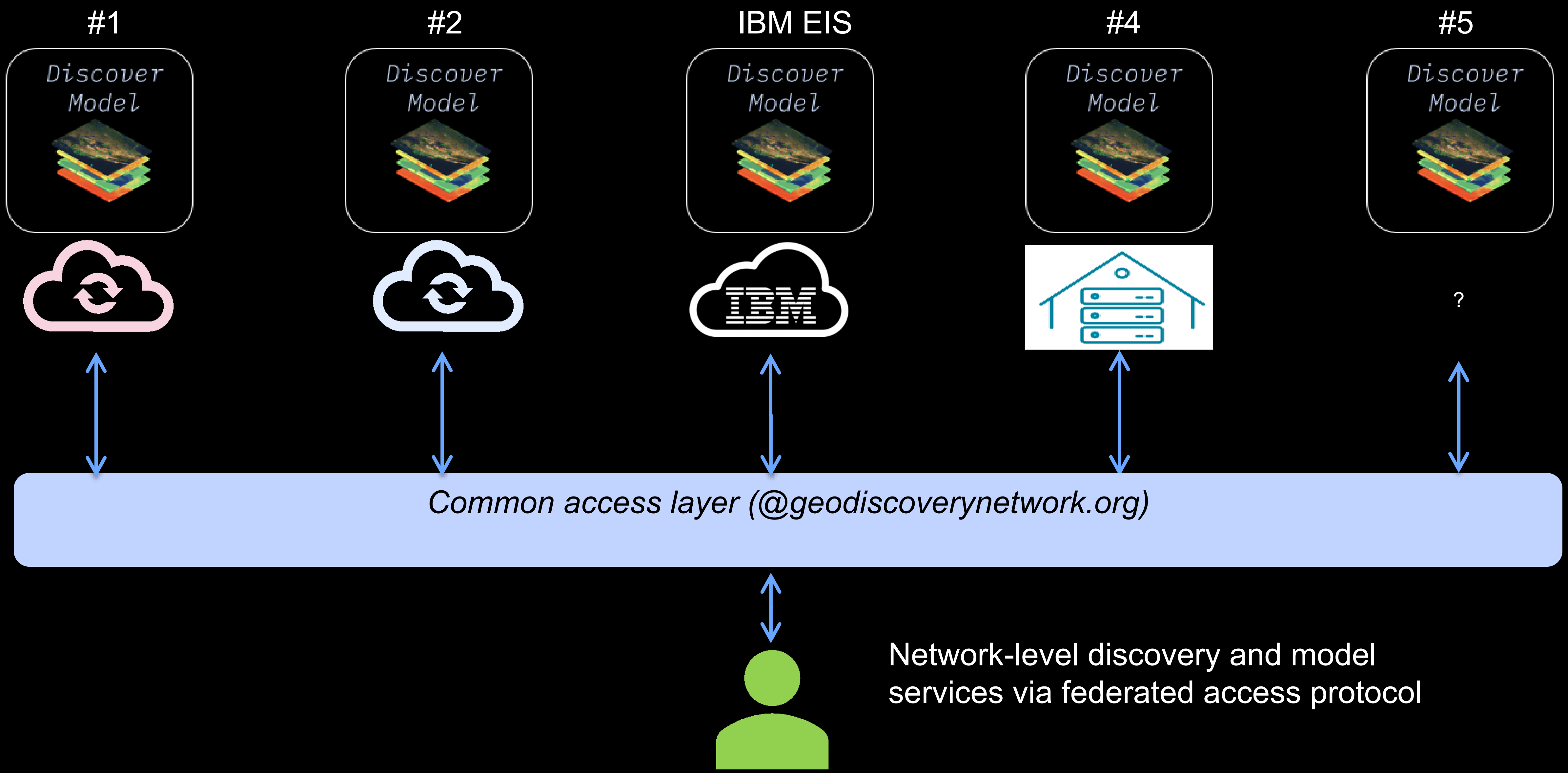


Geospatial Discovery Network –Phase 1

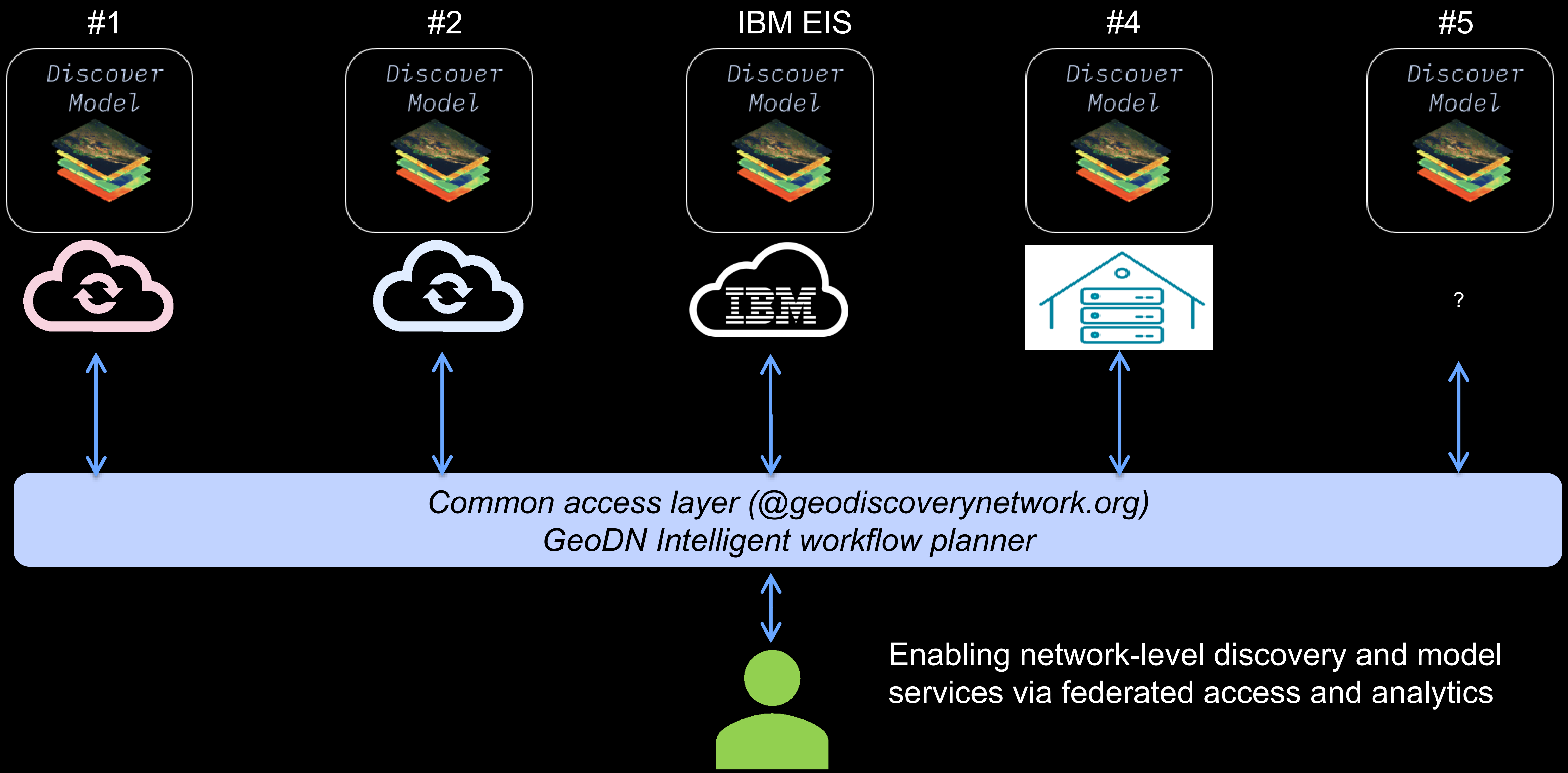


Node-level discovery and model services via node-level access protocol

Geospatial Discovery Network –Phase 2



Geospatial Discovery Network –Phase 3



Enabling network-level discovery and model services via federated access and analytics

Thank you