Radiant MLHub:
A Repository for Machine Learning Ready Geospatial Training Data

HAMED ALEMOHAMMAD
CHIEF DATA SCIENTIST

Radiant Earth Foundation
EARTH IMAGERY FOR IMPACT

13 NOV 2019
LOCATION POWERS: DATA SCIENCE

Mission

Empowering organizations and individuals globally with open AI and EO data, standards and tools to address the world's most critical international development challenges.

Vision

A strengthened global development community that benefits from high-quality and trusted AI and EO resources for positive global impact.

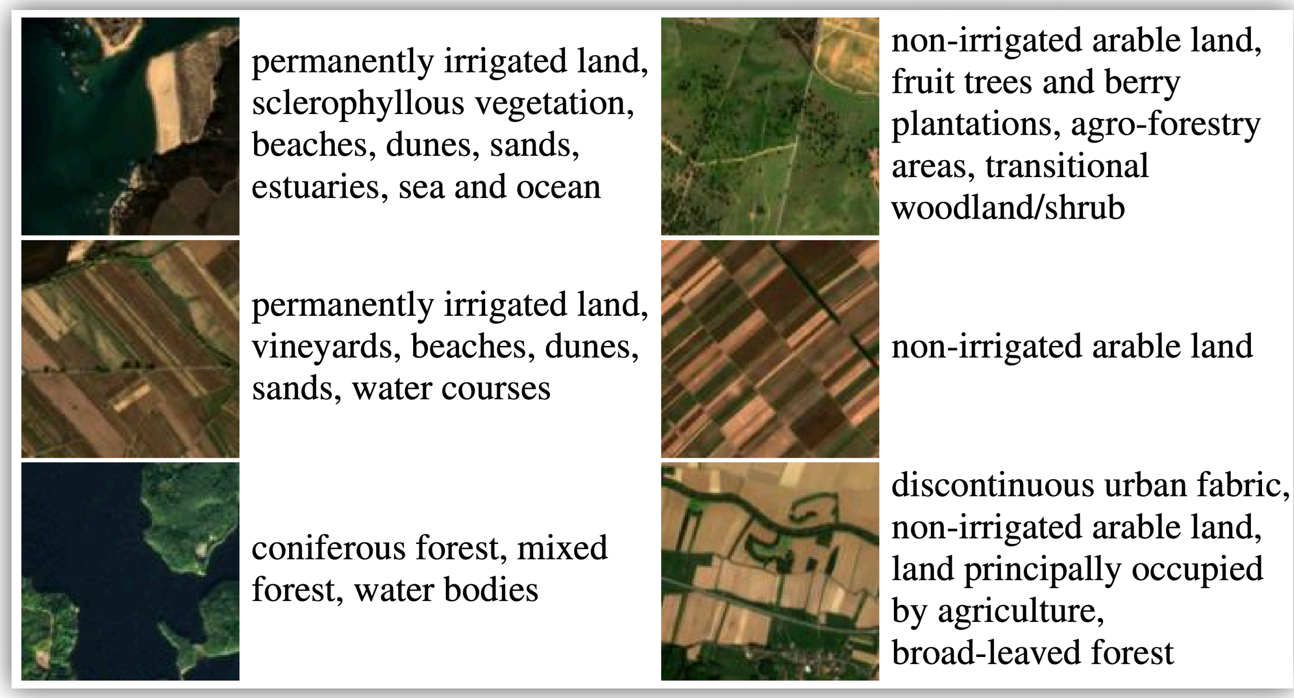
Motivation

Benchmark for Computer Vision

- 14 M annotated images including 1 M with object bounding boxes.
- 20 K categories of objects
- Open access
- Annual competition since 2010

BigEarthNet

- 590,326 Sentinel-2 image patches
- Each patch is a section of i) 120 × 120 pixels for 10m bands; ii) 60 × 60 pixels for 20m bands; and iii) 20 × 20 pixels for 60m bands.
- Each image patch is annotated by multiple land-cover classes
- 95% of images have at most 5 multi-labels.
- Images are selected across 10 EU countries.
- Data available for download on the web



SpaceNet

- A corpus of commercial satellite imagery and labeled training data to use for machine learning research.
- Training data available for open access on AWS

AOI	Area of Raster (Sq. Km)	Building Labels (Polygons)	Road Labels (LineString)
Rio	2,544	4,082,529	N/A
Vegas	216	151,367	3685 km
Paris	1,030	23,816	425 km
Shanghai	1,000	92,015	3537 km
Khartoum	765	710,960	1030 km
Atlanta	655 x 27	126,747	3000 km

Challenges in Geospatial ML

Geospatial Training Data Catalogs:

- Lack of Geo-Diversity
- Scarce data sources
- Data Accessibility
- Inter-Operability
- Machine learning-readiness

Result of Gaps in Training Data Catalogs:

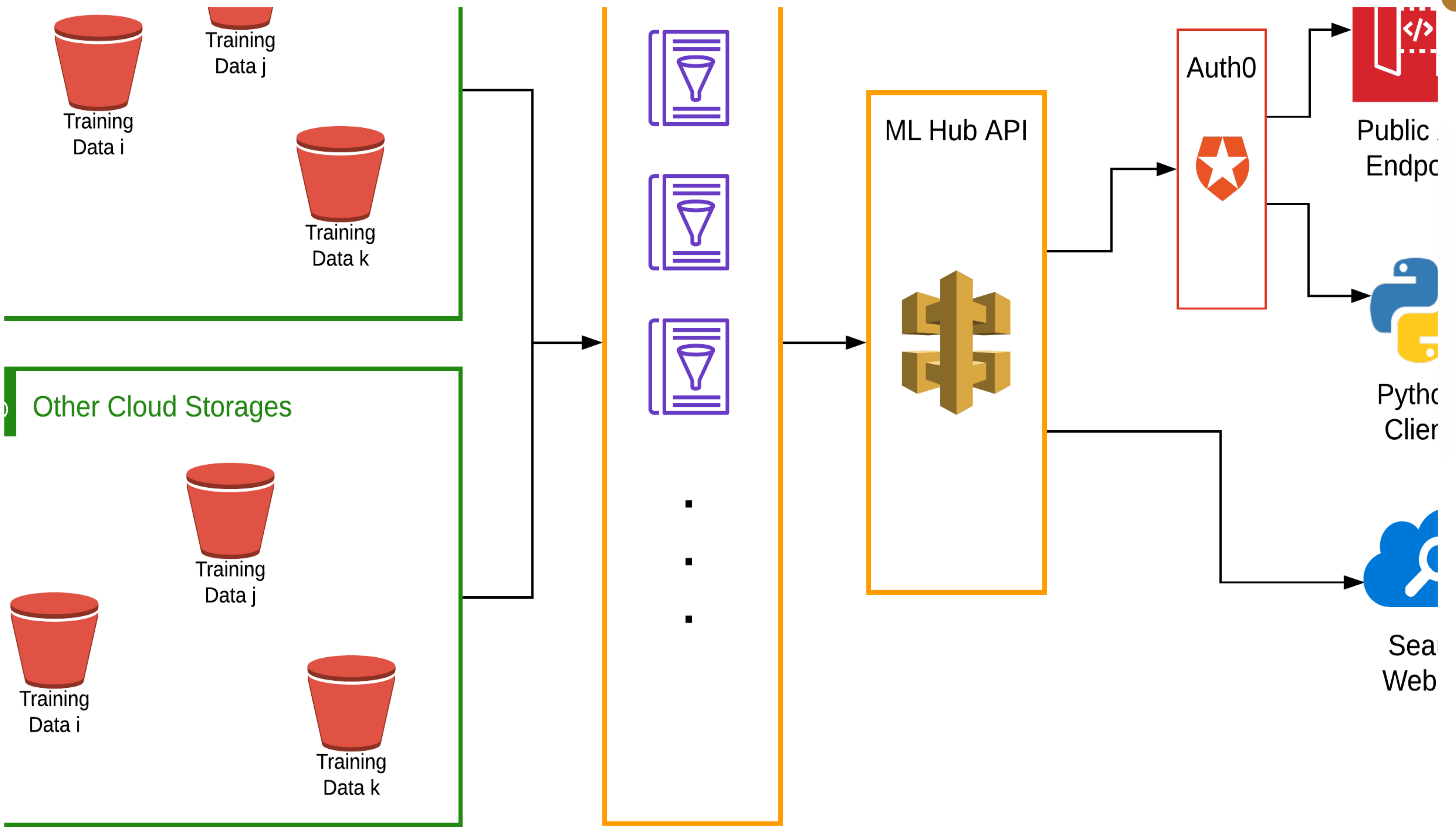
- Biased or incorrect results
- Inability to capture wide range of possible outcomes in space and time

ML Commons for Earth Observation

Radiant MLHub
Earth Imagery for Impact

Hub	Community	Educations
• EO training datasets • ML Models • High impact competitions • Image annotation + ground-referencing	• Convenings to develop standards around ML on EO • Interoperability of datasets • Technical Working Groups • White Papers	• EO market information • Best practices on use of ML and EO • Speaking engagements • Media outreach

Radiant ML Hub



An Open Source Standard

SpatioTemporal Asset Catalog (STAC)

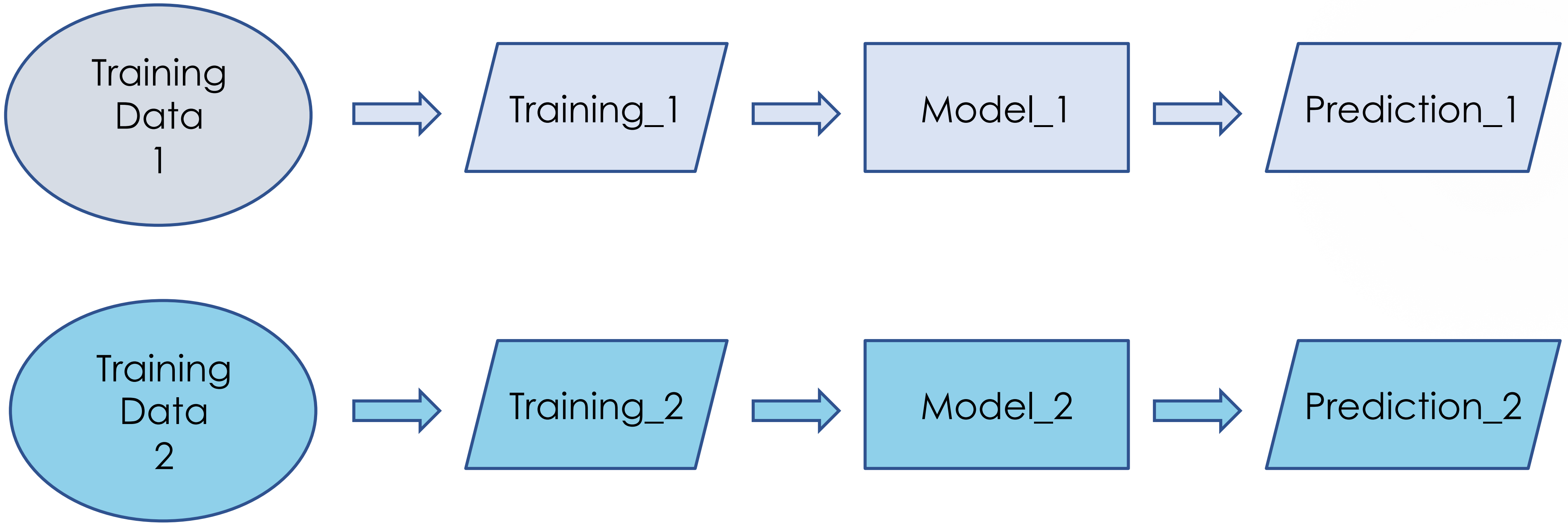
An open specification to increase the interoperability of searching for geospatial data.

STAC
SpatioTemporal
Asset Catalog

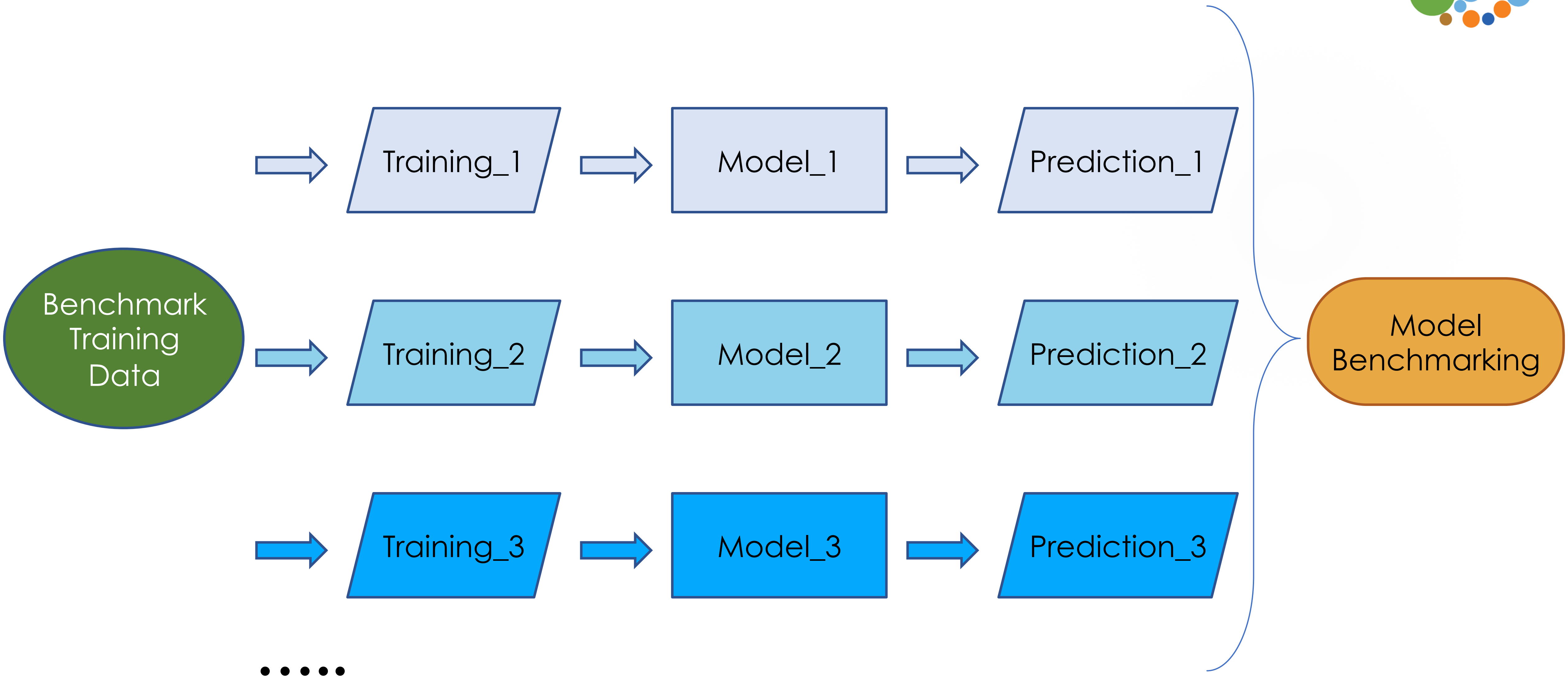
A Community:

MAXAR (Digital Globe), Azavea, Google Earth Engine, Harris, CBERS, Development Seed, USGS, Geoscience Australia, Planet, Astraea, Element-84, Radiant Earth Foundation

Existing Workflows



Ideal Workflow



How will it look like in practice?

```
In [1]: import mlhubearth as mle

In [2]: demo_bbox = [
                     (37.176361083984375, -6.815989239380505),
                     (37.24365234375, -6.7559877474853325)
                    ]

In [3]: td_id = mle.datasets(keywords = "maize", bbox = demo_bbox)

        {'data provider': 'Radiant Earth Foundation',
         'docs': 'www.mlhub.earth/datasets/tanzania_crop_2017',
         'id': '87472972523egb7365288157394uyhs6352527e',
         'license': 'CC BY',
         'name': 'Tanzania Crop Type Dataset 2017',
         'version': '1.0'}

In [4]: (x_train, y_train) = td_id.load_data()

In [ ]:
```

Thanks!

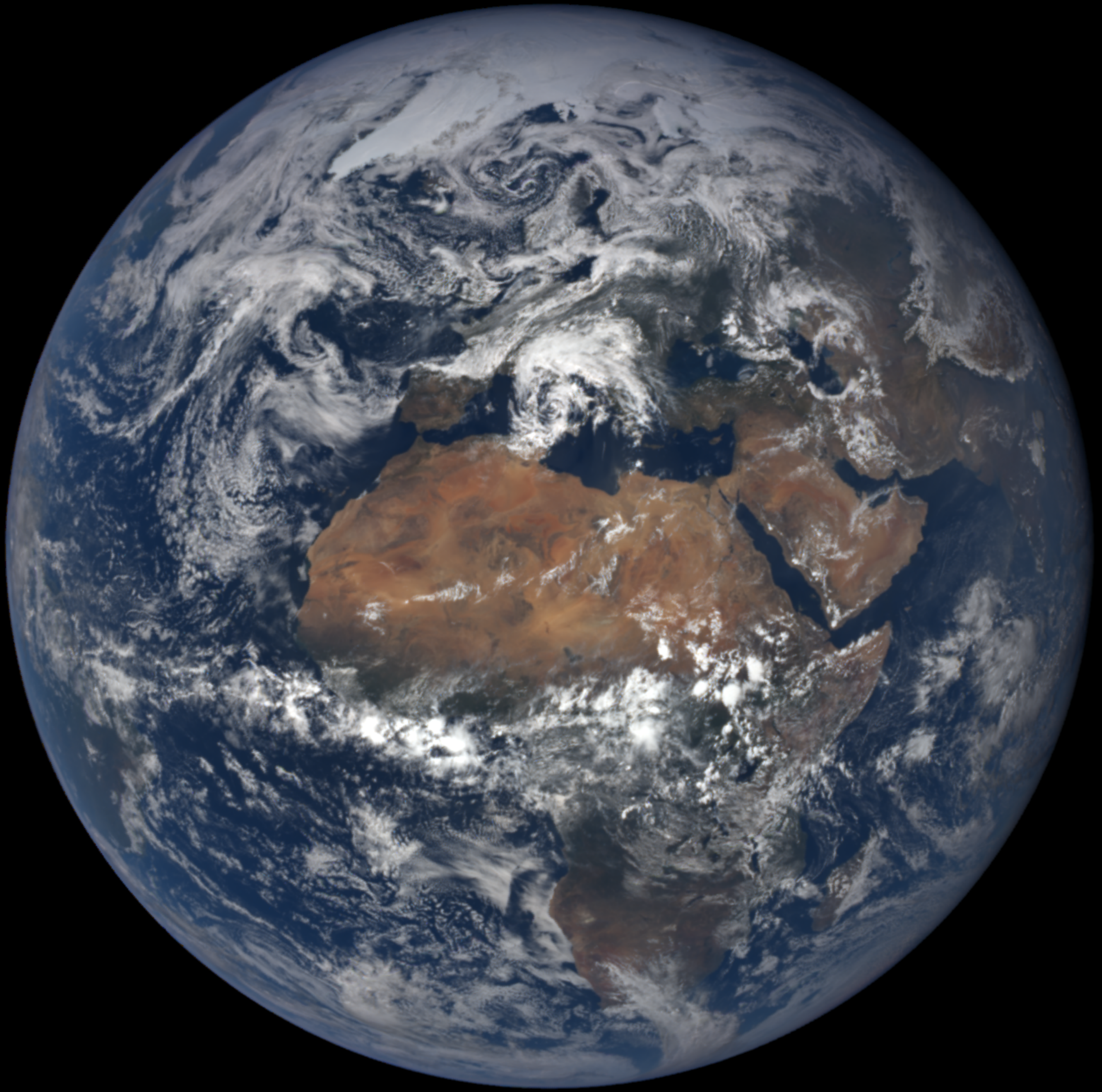
hello@radiant.earth

www.radiant.earth

OurRadiantEarth

github.com/radiantearth

MLHub Slack Channel:
bit.ly/MLHubSlack



Funders

OMIDYAR NETWORK

BILL & MELINDA GATES foundation

SCHMIDT **FUTURES**

Patrick J. McGovern FOUNDATION

NASA