

AgReFed

Data-Harvester

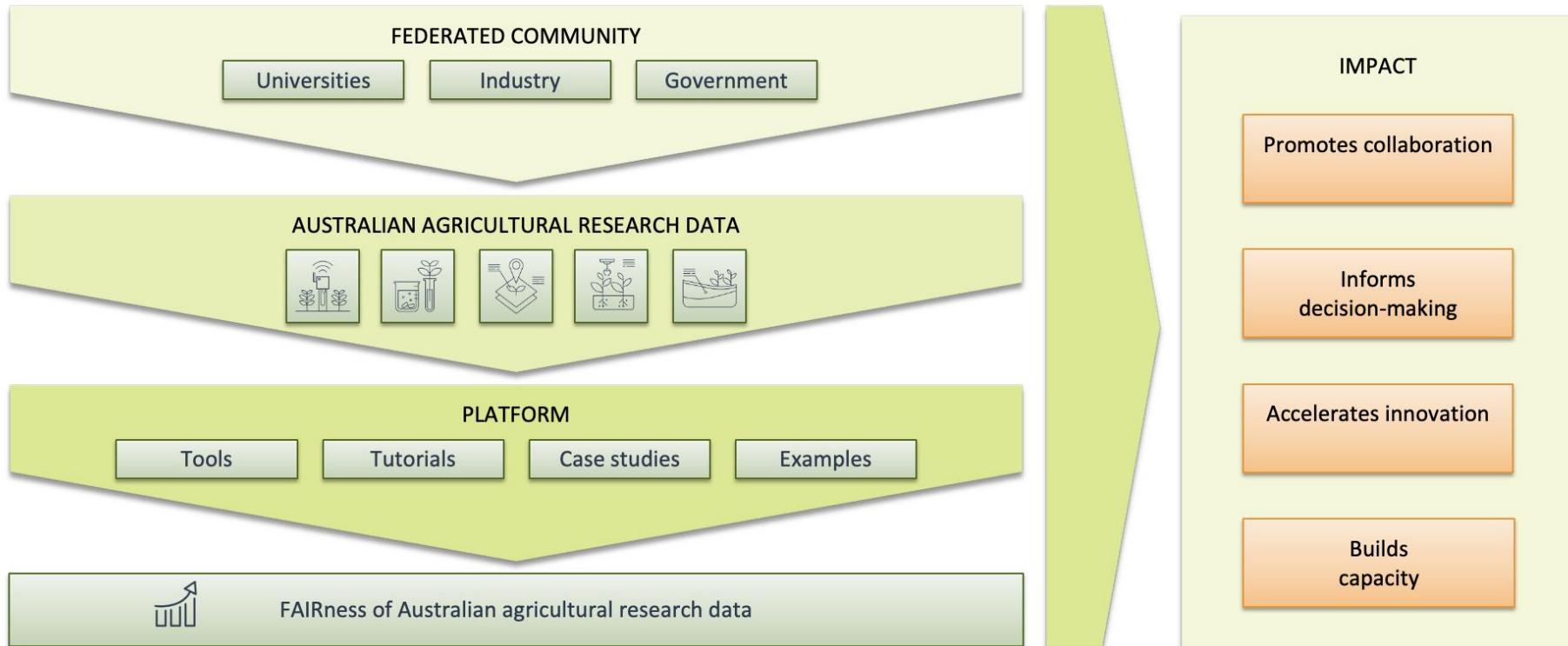
Jumpstarting research and reusable Machine Learning workflows

Sebastian Haan | Januar Harianto | Nathaniel Butterworth | Thomas Bishop
Sydney Informatics Hub, The University of Sydney

Our shared mission

“...enable **Findable, Accessible, Interoperable and Reusable (FAIR)** agricultural data to accelerate innovation and increase the profitability and sustainability of Australian agriculture”

AgReFed – Overview

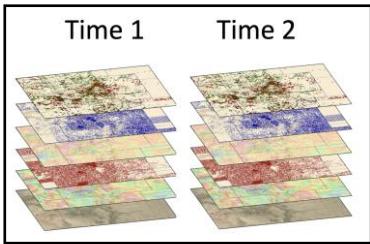


AgReFed Data-Harvester

Enable researchers to:



Retrieve — automatically access geospatial and soil data sources, minimal handling of individual APIs (software takes care of it)



Process — filter, mask, reduce and convert data

Bulk_Density_0-5cm	Clay_0-5cm	monthly_rain_Total	max_temp_Me
1.368779	27.214527	47.000000	37.50
1.362662	31.956041	47.399902	37.29
1.362662	31.956041	47.399902	37.29
1.360451	32.675858	35.899902	37.60
1.334362	35.097813	35.899902	37.60

Analyse — download data as GeoTIFF and/or tidy data frames for use in additional modelling and machine learning workflows

AgReFed Data-Harvester – Data Layers

Australian:

Soil Data	Climate Data	Satellite Data	Terrain	Radiometrics
<ul style="list-style-type: none">• Organic Carbon• Depth of Soil• Clay• pH• ...	<ul style="list-style-type: none">• Rainfall• Temperature• Evaporation• Drought• ...	<ul style="list-style-type: none">• Sentinel, Landsat• Vegetation• Surface Reflectance• Water• ...	<ul style="list-style-type: none">• Elevation• Slope, Aspect• Landscape	<ul style="list-style-type: none">• Gamma• Thorium, Uranium• ...

Bonus:



Google Earth Engine Integration

Spatial-Temporal Processing



SPATIAL

- Point data extraction/spatial buffering from images
- Raster alignment and transformation
- Conversion of coordinate reference systems
- Polygon extraction
- Masking (e.g. cloud-cover)

TEMPORAL

- Temporal aggregation and stats
- Temporal interpolation
- Temporal “buffer” kernel functions

AgReFed Data-Harvester – Automation

Settings

YAML
configuration
file

```
#####
#          #
#      Settings Data-Harvester      #
#          #
# This is the settings file for the notebook.      #
# For more information on the available Data source #
# and layers, see README and Data_Overview.md      #
#####

##### Input and Output Specifications #####
# Input File:
infile: ./testdata/Pointdata_Llara.csv
# Output Path:
outpath: ../../dataresults/
# Headername of Latitude in input file:
colname_lat: Lat
# Headername of Longitude in input file:
colname_lng: Long

##### Settings for Spatial and Temporal Specifications #####
# Bounding Box as [lng_min, lat_min, lng_max, lat_max]:
# if None provided, bounding box will be inferred from input coordinates
target_bbox: ''
# Select years:
target_dates:
- 2021
-- 2020
# Spatial Resolution [arcsec]:
target_res: 6.0
```

**Input file and
feature names**

**Spatial and temporal
settings: bbox,
dates, resolution, ...**

Settings

YAML configuration file

```
##### Data Selection #####
target_sources:
  # Satellite data from Digital Earth Australia (DEA) Geoscience Earth Observations
  DEA:
    - landsat_barest_earth

  # National Digital Elevation Model (DEM) 1 Second
  DEM:
    - DEM
    #- Slope
    #- Aspect

  # Landscape from SLGA
  Landscape:
    - Slope
    - Aspect
    - Relief_300m

  # Radiometric Data
  Radiometric:
    - radmap2019_grid_dose_terr_awags_rad_2019
    - radmap2019_grid_dose_terr_filtered_awags_rad_2019

  # SILO Climate Database
  # options for aggregation: Median, Mean, Total
  SILO:
    max_temp:
      - median
    min_temp:
      - median
    monthly_rain:
      - sum

  # Soil data from Soil and Landscape Grid of Australia (SLGA)
  # Depth options: 0-5cm, 5-15cm, 15-30cm, 30-60cm, 60-100cm, 100-200cm
  SLGA:
    Bulk_Density:
      - 0-5cm
```

Data layer selections:
Landsat, Terrain, SLGA,..

Temporal aggregation methods:
monthly, weekly,
daily mean,...

AgReFed Data-Harvester – Automation

Settings - Edit YAML using interactive widget

New Settings Load Settings

▼ Input and Output Specifications

Input File: No selection

Output Path:

Headername of Longitude:

Headername of Latitude:

▶ Settings for Spatial and Temporal Specifications

▶ SLGA Data Selection

▶ SILO Data Selection

▶ DEA Data Selection

▶ DEM Data Selection

▶ Radiometrics Data Selection

▶ Landscape Data Selection

AgReFed Data-Harvester – Automation

Settings - Edit YAML using interactive widget

New Settings Load Settings

▶ Input and Output Specifications

▶ Settings for Spatial and Temporal Specifications

▶ SLGA Data Selection

▼ SILO Data Selection

<input type="checkbox"/> daily_rain	Temporal aggregation:	mean median	Daily rainfall, mm
<input type="checkbox"/> monthly_rain	Temporal aggregation:	mean median	Monthly rainfall, mm
<input type="checkbox"/> max_temp	Temporal aggregation:	mean median	Maximum temperature, degrees Celsius
<input type="checkbox"/> min_temp	Temporal aggregation:	mean median	Minimum temperature, degrees Celsius
<input type="checkbox"/> vp	Temporal aggregation:	mean median	Vapour pressure, hPa
<input type="checkbox"/> vp_deficit	Temporal aggregation:	mean median	Vapour pressure deficit, hPa
<input type="checkbox"/> evap_pan	Temporal aggregation:	mean median	Class A pan evaporation, mm

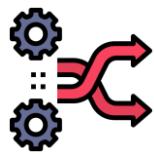
Workflow

harvest()

Validate



API Access



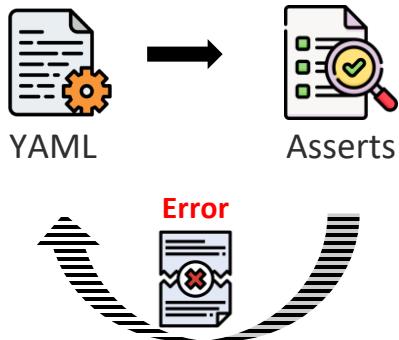
Process



Finalise

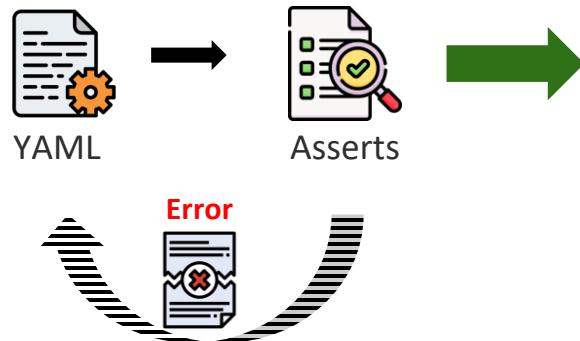


Validate

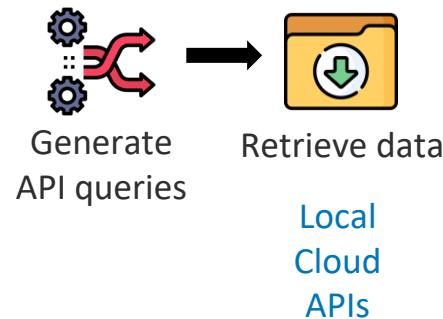


1. Validate input variables before pre-processing

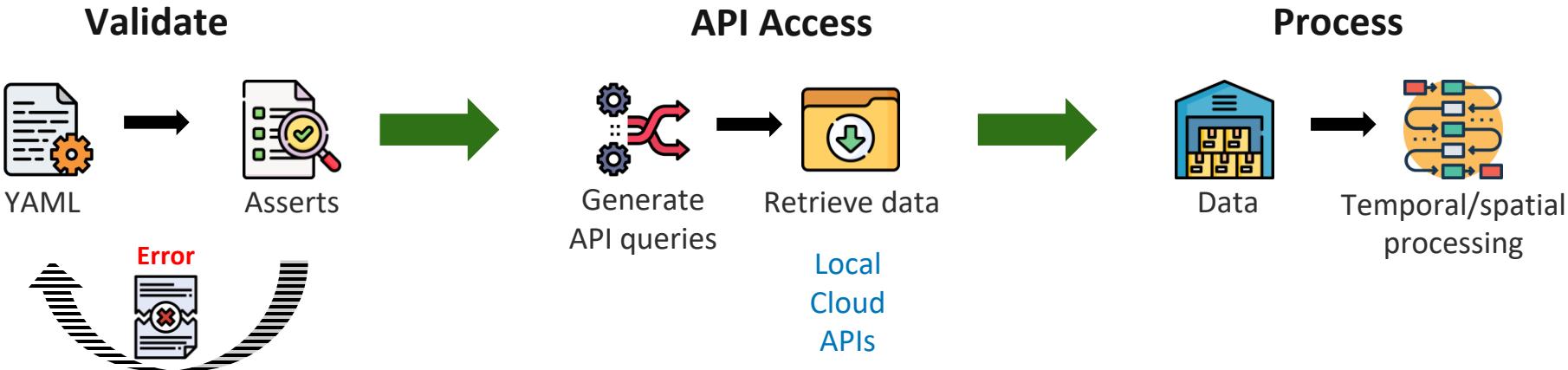
Validate



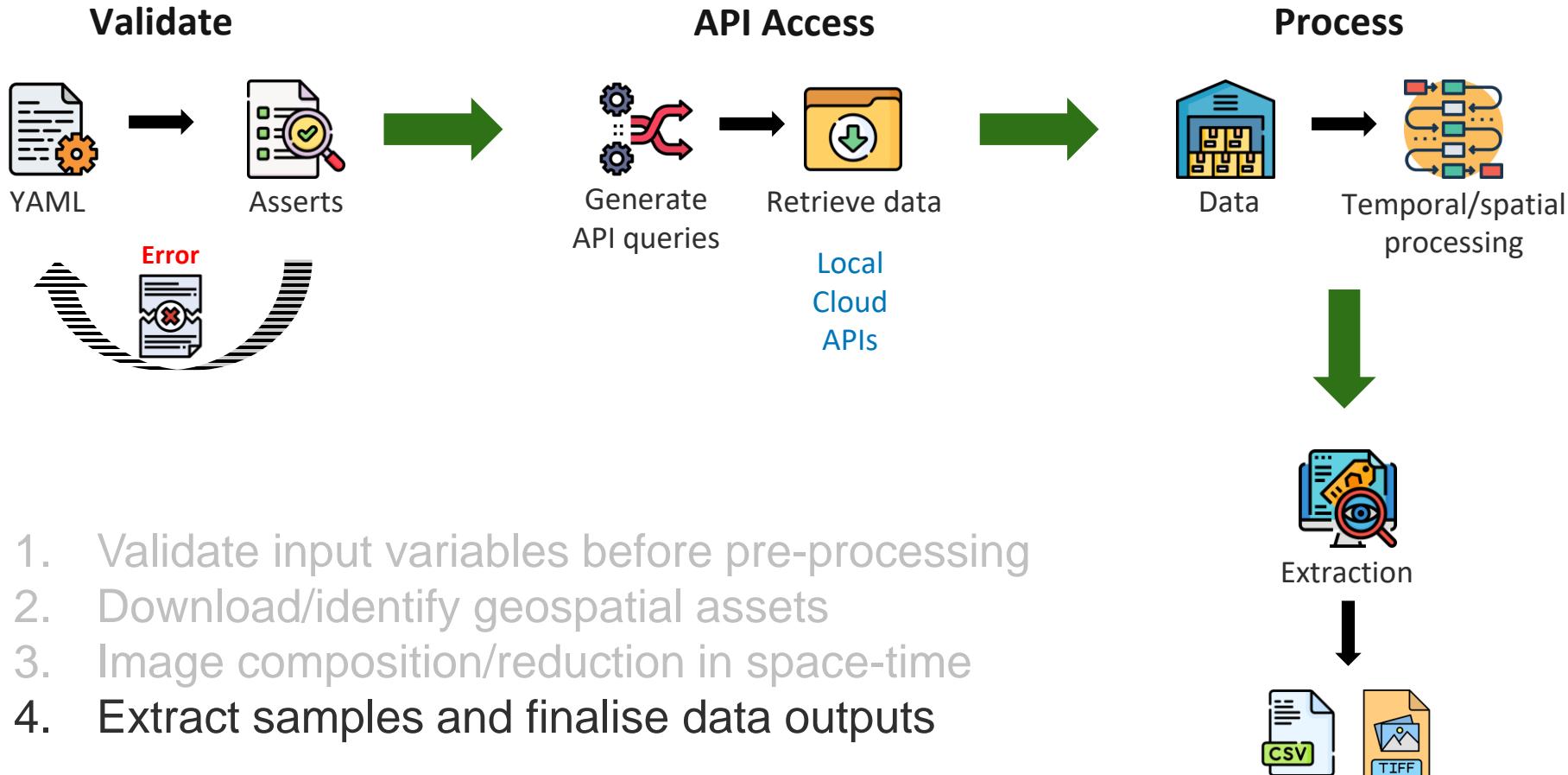
API Access



1. Validate input variables before pre-processing
2. Download/identify geospatial assets



1. Validate input variables before pre-processing
2. Download/identify geospatial assets
3. Image composition/reduction in space-time

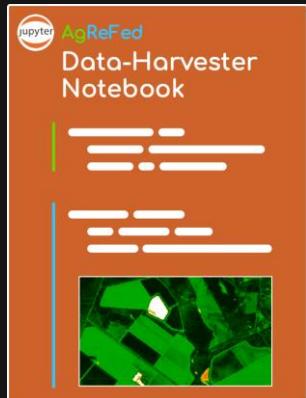


Three (3) products

(so far...)



Data-Harvester



APIs and Python Frontend



eeharvest



Google Earth Engine API



dataharvester



R package



dataharvester

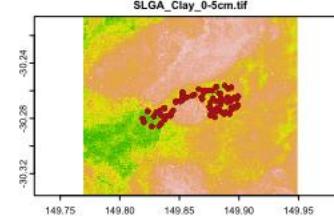
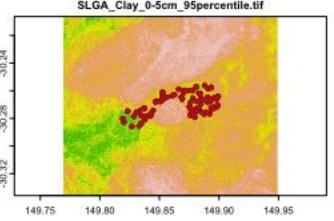
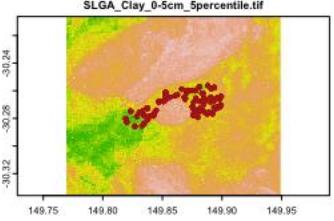
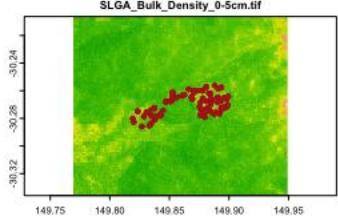
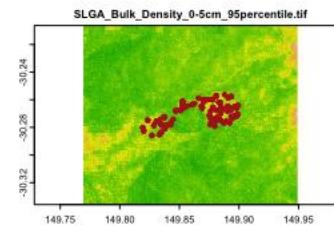
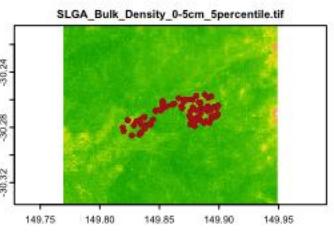
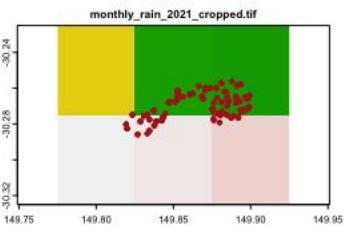
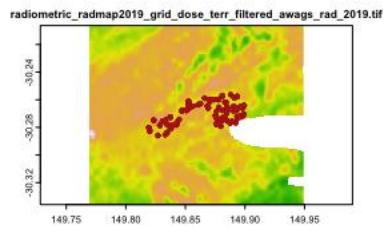
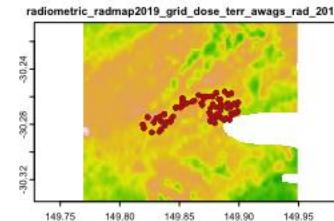
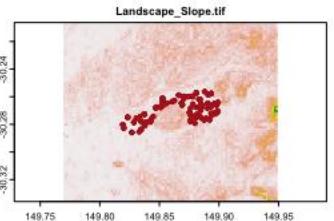
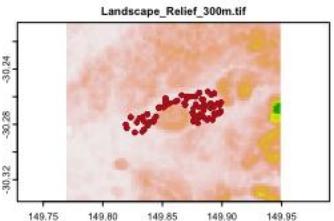
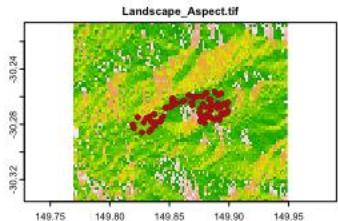
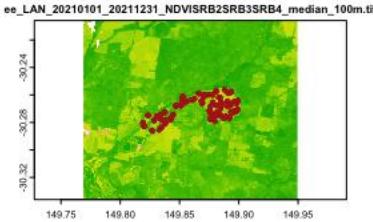
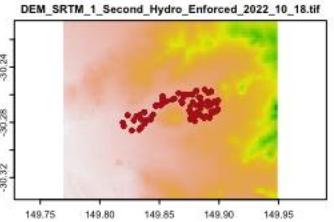
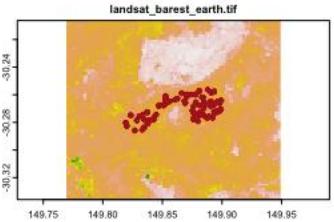
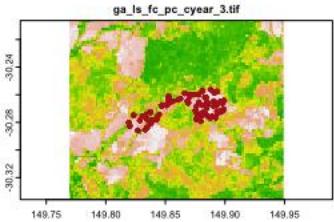


R package



Restarting R session ...

> |





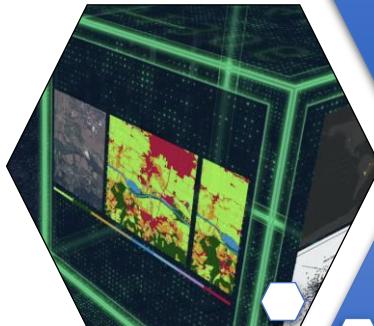
		Longitude	Latitude	geometry	ee_LA... ¹	lands... ²	ga_ls... ³	DEM	lands... ⁴	lands... ⁵	lands... ⁶	radma... ⁷	radma... ⁸	month... ⁹	Bulk_... ¹⁰	Clay_... ¹¹
1	0	150.	-30.3	POINT (149.85268...	0.693	1059	4	245.	1.05	209.	10.5	33.2	33.0	47	1.37	27.2
2	1	150.	-30.3	POINT (149.85268...	0.693	1059	4	245.	1.05	209.	10.5	33.2	33.0	47	1.37	27.2
3	2	150.	-30.3	POINT (149.85268...	0.693	1059	4	245.	1.05	209.	10.5	33.2	33.0	47	1.37	27.2
4	3	150.	-30.3	POINT (149.85268...	0.693	1059	4	245.	1.05	209.	10.5	33.2	33.0	47	1.37	27.2
5	4	150.	-30.3	POINT (149.85268...	0.693	1059	4	245.	1.05	209.	10.5	33.2	33.0	47	1.37	27.2
6	5	150.	-30.3	POINT (149.88483...	0.222	1082	4	264.	1.00	280.	6.04	36.0	35.9	47.4	1.36	32.0
7	6	150.	-30.3	POINT (149.88483...	0.222	1082	4	264.	1.00	280.	6.04	36.0	35.9	47.4	1.36	32.0
8	7	150.	-30.3	POINT (149.88483...	0.222	1082	4	264.	1.00	280.	6.04	36.0	35.9	47.4	1.36	32.0
9	8	150.	-30.3	POINT (149.88483...	0.222	1082	4	264.	1.00	280.	6.04	36.0	35.9	47.4	1.36	32.0
10	9	150.	-30.3	POINT (149.88483...	0.222	1082	4	264.	1.00	280.	6.04	36.0	35.9	47.4	1.36	32.0



> |

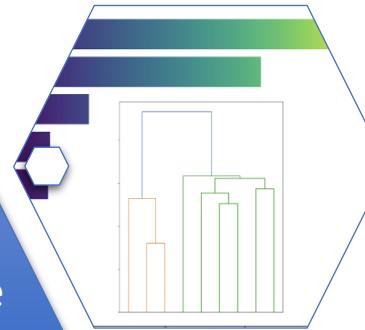
AgReFed Data Cube: machine learning

“Developing reusable workflows for machine learning to understand agriculture systems and their uncertainties.”

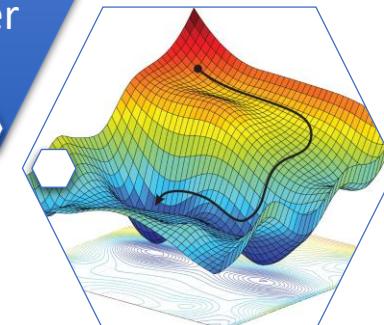


Probabilistic
Soil Mapping
and Data
Cubes

Auto-Feature
Importance
Analysis



Spatial-
Temporal
Models under
Uncertainty



AgReFed Data-Harvester: Summary

Flexibility by modular design.



High-level access via notebooks: easily editable and readable, automated and reusable workflow to retrieve and transform data into ready-made dataframes



"Under-the-hood" Python modules: for experienced users and open-source developers



API encapsulation and extensions: one module per API handler; enables easy addition of new APIs to Data Harvester with example templates.



Flexible deployment: any cloud infrastructure or local machine



<https://sydney-informatics-hub.github.io/geodata-harvester/>



Nov. 03



AgReFed Data Harvester Workshop

Learn to use the AgReFed Data-Harvester platform to rapidly retrieve geospatial data from a wide range of open online sources.



THE UNIVERSITY OF
SYDNEY Sydney Informatics Hub

Point of Contact:

Januar Harianto

januar.harianto@sydney.edu.au

Sebastian Haan

sebastian.haan@sydney.edu.au

Questions?

Discussion Points

- Best practices in data sharing to power future reusable workflows and modern data-pipelines
- Preferred API methods for spatial-temporal data exchange
- Metadata and licensing for data pairing and derived data
- ...



<https://sydney-informatics-hub.github.io/geodata-harvester/>