

# LEVERAGING AGRICULTURAL DATA

## THE ROLE OF STANDARDS



Platform for  
Big Data  
in Agriculture

**M. Devare**

Senior Research Fellow, IFPRI  
Module Lead, CGIAR Big Data Platform

# A FUTURE FARM

Feeding the future. Byte by byte.

Standard A

## SURVEY DRONES

Aerial drones survey the field to map crop health and yield, and soil variation. **Cloud storage** enables faster data movement, analysis and exchange, **increasing computational capacity** while decreasing costs.



## DATA-DRIVEN SOLUTIONS

Find data  
Interpret  
Aggregate

Visualize  
Analyze

Standard B

## FARM TO CLOUD

Farms generate vast amounts of **rich and varied data**, which can now be stored, processed, and shared in the cloud, to be delivered **back to farms as important advisory services**.



Standard C

## UBER AGRIBOTS

**Peer-to-farmer services** allow smallholder farmers access to cost-effective cutting edge technology such as agribots that can tend to or harvest crops, and provide precise application of fertiliser and water, **saving economic and environmental costs**.



## LIVESTOCK LIVE

**Decreasing costs** of small devices, such as mobile phones and sensors, **are increasing access** to get rural farmers - and their livestock - online and benefiting from data exchange.



Platform for  
Big Data  
in Agriculture

# Harnessing digital in agriculture: Data must be Open and FAIR



## Findable

Metadata and data should be findable for both humans and computers

## Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

F

A

I

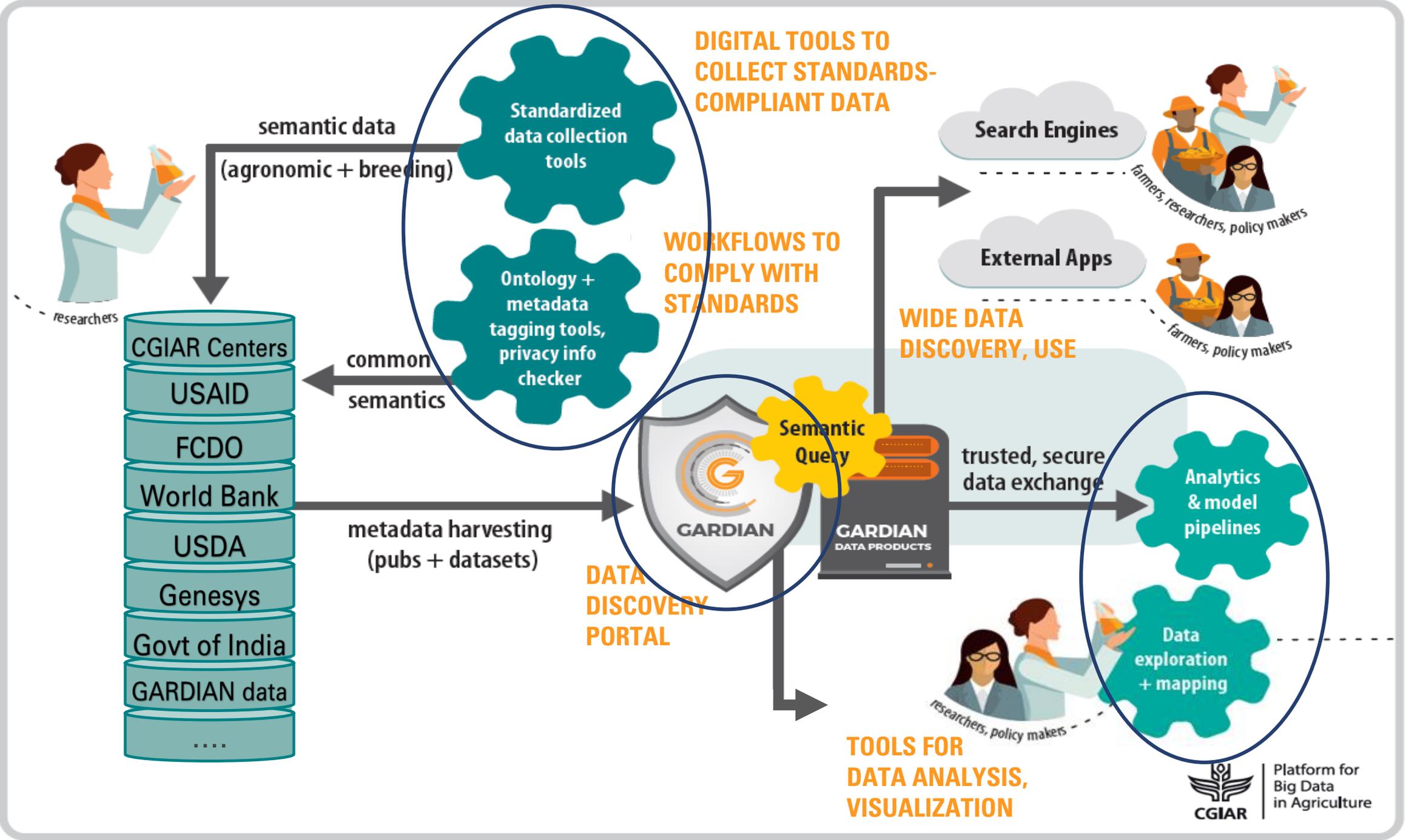
R

## Accessible

Once found, users need to know how the data can be accessed

## Reusable

The goal of FAIR is to optimise data reuse via comprehensive well-described metadata



**DIGITAL TOOLS TO COLLECT STANDARDS-COMPLIANT DATA**

**WORKFLOWS TO COMPLY WITH STANDARDS**

**WIDE DATA DISCOVERY, USE**

**DATA DISCOVERY PORTAL**

**TOOLS FOR DATA ANALYSIS, VISUALIZATION**

- CGIAR Centers
- USAID
- FCDO
- World Bank
- USDA
- Genesys
- Govt of India
- GARDIAN data
- ....



**GARDIAN**

<https://gardian.bigdata.cgiar.org>

Global Agricultural Research Data Innovation Acceleration Network

crop fertilizer Africa

Welcome to the GARDIAN ecosystem

Are you new to GARDIAN?  
Learn more about the GARDIAN Ecosystem >



**191,570**  
PUBLICATIONS



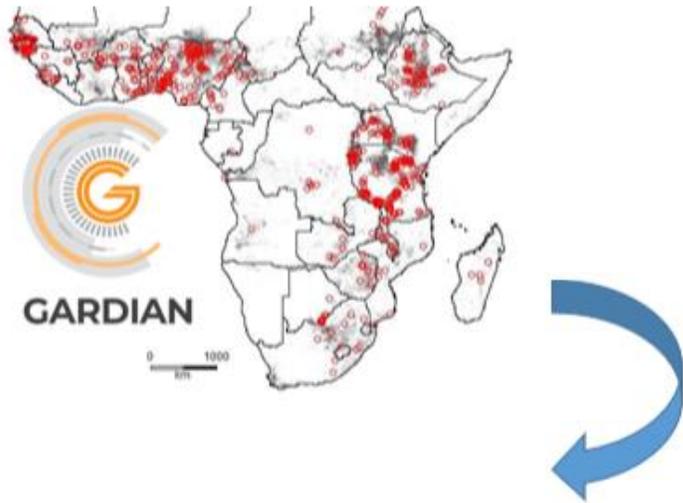
**38,735**  
DATASETS



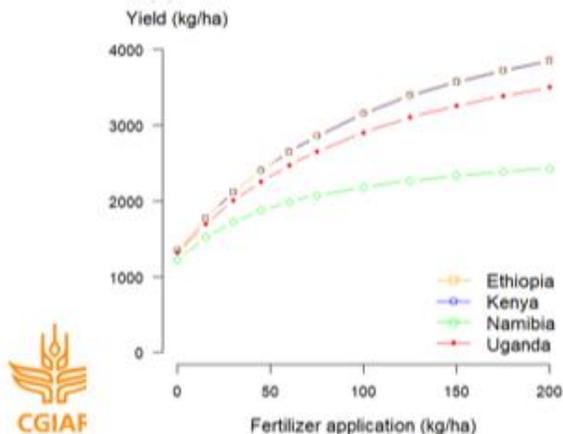
**DATA MANAGEMENT  
TOOLKIT**

# OK...so what??

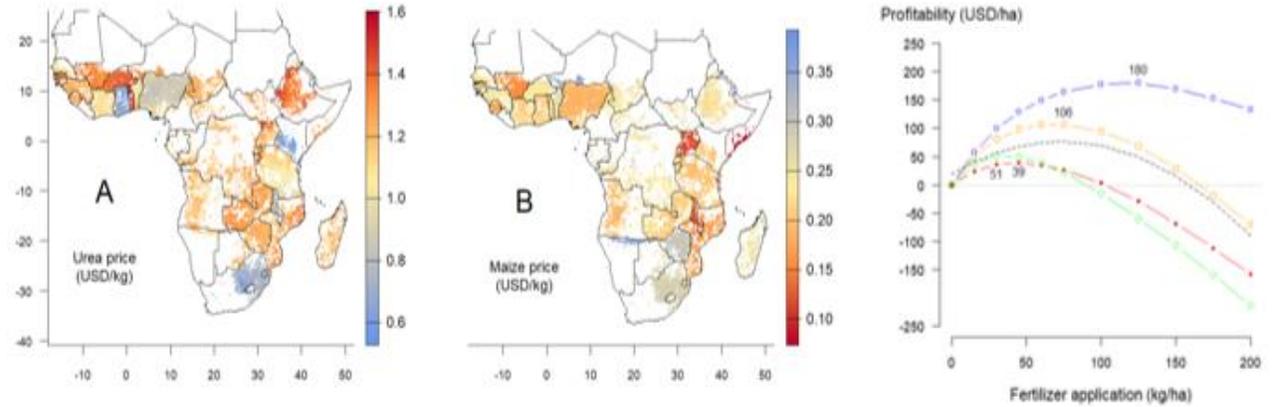
1) Find fertilizer response data from 760 locations in Africa - GARDIAN



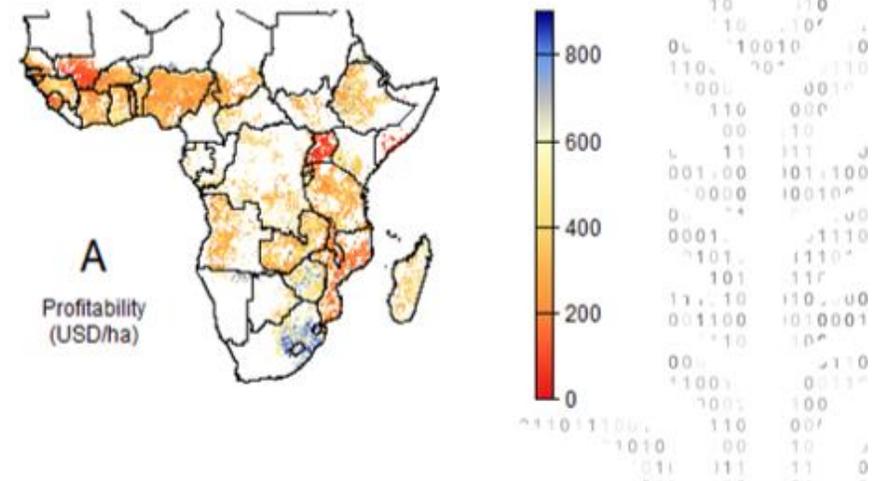
2) Develop machine learning model



3) With price data, compute local profitability of fertilizer use



4) Assess profitability of fertilizer use



Bonilla (CIAT), Chamberlin (CIMMYT), Hijmans (UC Davis)

# Interoperable?

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Year	NRD	GSRF	Evaporat	WindS	Radiation	MinT	MaxT	MinRH	MaxRH	Sunhr
3	1982	53	516.2	4.2	4.2	17.8	22.6	30.1	63.5	97.3	5.4
4	1983	52	661.3	4.4	4.6	17.2	22.7	30.7	61.1	98.0	5.5
5	1984	56	636.6	4.2	4.9	18.7	22.6	30.5	61.5	96.7	6.2
6	1985	51	827.4	4.0	4.3	17.9	22.7	29.8	61.1	95.0	5.7
7	1986	45	658	4.1	4.6	16.7	22.6	30.0	61.5	95.1	5.7
8	1987	44	626.6	4.4	4.5	17.9	23.7	31.6	60.1	95.0	6.1
9	1988	49	770.5	4.0	4.0	15.1	22.5	29.5	64.8	96.7	5.5
10	1989	52	652	4.3	4.2	18.4	22.7	30.2	60.7	97.7	5.9
11	1990	66	645.6	4.0	4.4	18.3	22.2	30.4	63.9	97.2	5.5
12	1991	62	913.1	4.2	4.3	15.1	22.1	30.9	62.2	96.9	5.5
13	1992	57	638	3.9	4.3	15.4	22.0	30.7	61.0	96.7	4.8
14	1993	48	531.4	3.8	4.0	13.8	21.7	30.9	63.5	97.6	5.9
15											
16											
17	Year	TotalRF	Onset	Cessation	Rainydays	Drydays	MI1S	MI2S	MI3S	MinT	MaxT
18	1982	728.2	15	92	57	140	136.3	267.4	324.5	23.0	31.0
19	1983	700.6	57	92	39	116	162.1	493.7	44.8	23.1	31.5
20	1984	973.9	42	92	56	114	154.5	462.9	356.5	22.7	31.0
21	1985	1171	40	92	58	114	239.4	329.3	602.3	23.1	30.6
22	1986	785.9	34	92	39	139	138.1	430.5	217.3	22.9	31.2
23	1987	1079.9	37	92	55	120	206.3	350.5	523.1	23.8	32.3
24	1988	827.2	63	92	47	102	236.6	515.7	74.9	22.4	30.1
25	1989	890	55	92	57	100	118.8	399.5	371.7	23.0	30.8
26	1990	719.1	59	92	51	102	268.1	299	152	22.4	31.3

# Interoperable?

```
# Link: http://gardian.bigdata.cgiar.org/dataset.php?id=1990
# Description: Drought is a key maize (Zea mays L.) producti
# Fourteen, fifteen, and twenty-five extra-early maturing ma
# Striga resistance and drought and low soil N tolerance, we
# developed from 1995 to 2000 (Period 1), 2001 to 2006 (Peri
# respectively.

# Extra material:
GainsinGrainYieldofExtra-earlyMaizeduringThreeBreedingPeriod
#####
library(raster)

path <- 'Gardian/Gains_grain_yield/1_Gardian_Gains_Grain_yie

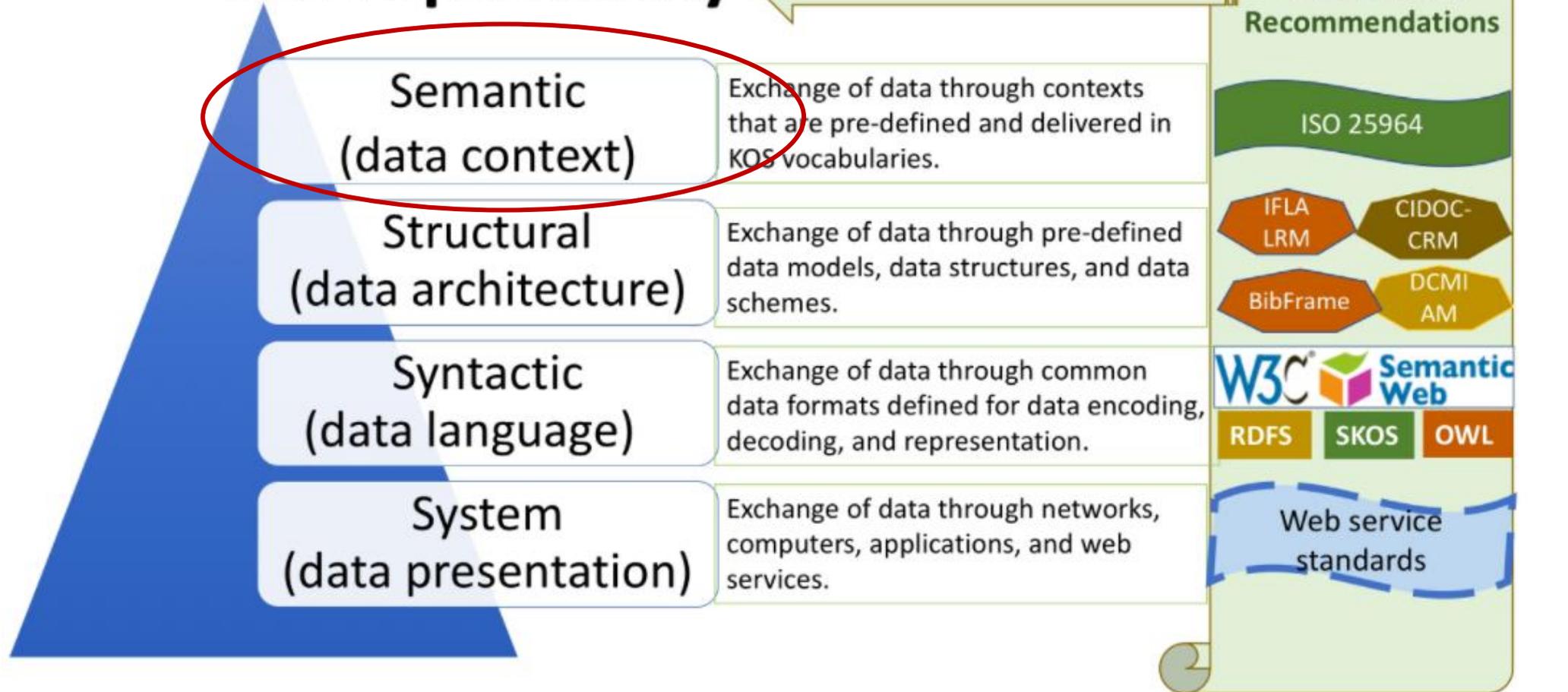
df <- read.csv(file.path(path, '/20181029aao_combd_ba15150_fi
14)']
colnames(df) <- c('country', 'town_village', 'treatment_code',
, 'crop_variety', 'yield_kg_ha')

df$country <- 'Nigeria'
df$country <- as.factor(df$country)
df <- aggregate(.~town_village, data=df, quantile, probs=0.75)

df$yield_kg_ha <- round(df$yield_kg_ha)
```

# The holy grail...

## Interoperability



# Can I generate interoperable agronomic data at collection?

## Agronomy Field Information Management System (AgroFIMS)

- Generates standardized field books to collect agronomic data that is born FAIR



### Fieldbook design

via ontology-based variables, terminology, and units in modules representing typical cycle of operations in agronomic trials



### Digital data collection

with KDSmart, ODK or Field Book mobile apps



### Data analysis

via AgroFIMS statistical scripts (R-based) and reports



### Data archiving

through easy upload to institutional repository

<https://agrofims.org>

Home

# AgroFIMS

Agronomy Field Management System



*"If tool builders do their job well it may not be necessary for researchers to think about ontologies at all."*

~C. Parr, May 2020

The Agronomy Field Information Management System (AgroFIMS) allows users to create fieldbooks to collect agronomic data that is already tied to a metadata standard (the CG Core Metadata Schema, aligned with the standard Dublin Core), and semantic standards like the Agronomy Ontology (AgrO), generating data that is Findable, Accessible, Interoperable, and Reusable (FAIR) at collection.

AgroFIMS therefore standardizes data collection and description for easy aggregation and inter-linking across disparate datasets. By December 2020 you will be able to export the fieldbooks you create to the Android-based KDSmart or Field Book applications for data collection. Collected data can be imported back to AgroFIMS for statistical analysis and reports.

**Watch this [short video](#) that provides an overview of AgroFIMS!**

We have also made available the Agronomy API or AgrAPI, which is a RESTful web service API specification to enable access, exchange, and integration of agronomic data across systems and applications. You can use the API to retrieve research management information, experimental designs, crop measurements, and environmental variables from AgroFIMS. The AgrAPI blueprint can be implemented in different programming languages. Currently, AgrAPI is implemented in the R statistical programming language allowing you to analyze your data with the R statistical packages and graphics of your choice.

In 2021 AgroFIMS will allow you to set up agronomic survey questionnaires, for data collection via ODK. It will also allow easy upload of your "born FAIR" data to Dataverse repository platforms with Dublin Core-compliant metadata schemas.

Funding for AgroFIMS was provided by the Bill and Melinda Gates Foundation's Open Access, Open Data Initiative, and the CGIAR Platform for Big Data in Agriculture. AgroFIMS is under GPL license.

If you need support, please contact us at [agrofims@cgiar.org](mailto:agrofims@cgiar.org)



AgroFIMS

Experiments

Sites

Statistical analysis

Documentation

About

[agrofims@cgiar.org](mailto:agrofims@cgiar.org)

AgroFIMS v0.3.69 build 1

# Getting to data that is born FAIR



**AgroFIMS**

Experiments

Sites

Statistical analysis

Documentation

About

agrofims@cgiar.org

AgroFIMS v0.3.69 build 1

Home / Experiments / Manage experiment: **VVSG1612345240 (191)**

Experiment Personnel Fieldbooks

## Experiment details

Experiment ID  
VVSG1612345240

Experiment name \*  
On-farm cassava nutrient omission trial

Experiment project name \*  
African Cassava Agronomy Initiative (ACAI)

Experiment start date  
2021-02-03

Experiment end date  
2021-03-04

Type of experiment  
on-farm

Experiment objective  
1. To gain understanding of the local cassava production system and the influence of farm socio-economic and biophysical variability and historical and current management practices on cassava

## Funding Agency

Funding agency type  
Foundation

Funding agency name  
Bill and Melinda Gates Foundation

# Experiments description



- Site
- Crop
- Design**
- Fertilizer
- Management practices
- Crop Measurement
- Crop Phenology
- Weather
- Soil

## Experimental design ?

- Experiment design \***
- Completely Randomized Design (CRD)**
- Randomized Complete Block Design (RCBD)
- Full Factorial Completely Randomized Design
- Full Factorial Randomized Complete Block Design
- Split Plot Design
- Split-Split Plot Design

## Treatment description \*

Number of treatments

Repetitions

## Factor

Select one... 

Number of puddling passes

Tillage start date

Tillage end date

Tillage technique

Tillage depth

## Nutrient element type and amount

Type

Nitrogen 

Enter levels

0 

150 

75 

Unit

kg/ha 

Number of splits

3 



 Application details

Notes



## Factor

Nutrient element type and amount 

## Nutrient element type and amount



Parameter	Value
Experiment ID	VVSG1612345240
Experiment name	on-farm cassava NOT
Experiment project name	on-farm cassava nutrient omission trials
Experiment start date	2021-02-03
Experiment end date	2021-03-04
Type of experiment	
Experiment objective	
Grant number	
Grant Id	
Project management entity 1	CGIAR center
Project management center 1	IITA - International Institute of Tropical Agriculture
Project management contributor 1	
Experimental design abbreviation	crd
Experimental design	Completely Randomized Design (CRD)
Experimental unit	plot
Repetitions	2
Block	
Number of factors	4
Factor 1	Nutrient element type and amount1_f1
Factor 1-Levels	N0kg/ha, N150kg/ha, N75kg/ha
Factor 2	Nutrient element type and amount2_f2
Factor 2-Levels	P0kg/ha, P40kg/ha, P20kg/ha
Factor 3	Nutrient element type and amount3_f3
Factor 3-Levels	K0kg/ha, K180kg/ha, K90kg/ha
Factor 4	S-Ca-Mg-ZN-B4_f4
Factor 4-Levels	levels + units0kg/ha, levels + units16.6-10-10-5-5kg/ha

PLOT	ROW	COL	TREATMENT	1:Cassava_Soil_moisture_%_1	1:Cassava_Soil_moisture_%_2
1	1	1	1 N_75_kg/ha_/P_20_kg/ha_/K_90_kg/ha_/0_kg/ha		
2	1	2	2 N_0_kg/ha_/P_0_kg/ha_/K_0_kg/ha_/0_kg/ha		
3	1	3	3 N_150_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/16.6-10-10-5-5_kg/ha		
4	1	4	4 N_75_kg/ha_/P_20_kg/ha_/K_90_kg/ha_/0_kg/ha		
5	1	5	5 N_150_kg/ha_/P_0_kg/ha_/K_180_kg/ha_/0_kg/ha		
6	1	6	6 N_150_kg/ha_/P_0_kg/ha_/K_180_kg/ha_/0_kg/ha		
7	1	7	7 N_150_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/0_kg/ha		
8	2	7	7 N_0_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/0_kg/ha		
9	2	6	6 N_150_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/0_kg/ha		
10	2	5	5 N_150_kg/ha_/P_40_kg/ha_/K_0_kg/ha_/0_kg/ha		
11	2	4	4 N_0_kg/ha_/P_0_kg/ha_/K_0_kg/ha_/0_kg/ha		
12	2	3	3 N_150_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/16.6-10-10-5-5_kg/ha		
13	2	2	2 N_0_kg/ha_/P_40_kg/ha_/K_180_kg/ha_/0_kg/ha		
14	2	1	1 N_150_kg/ha_/P_40_kg/ha_/K_0_kg/ha_/0_kg/ha		

Measurement	TraitName	TraitUnit	NumberOfI	NumberOfV	Timing	TimingValue	SoilDepth	TraitLevel	TraitAlias	TraitDataTy	TraitValidation
Dry matter yield	Cassava_Storage roots_ kg/ha		1	1				Plot	Cas_DryMat	DECIMAL	0.00 <= x <= 100000.00
Plant height	Cassava_Whole plant_ P cm		1	5	Other	1st fertilization   2nd fertili		Plot	Cas_PlaHei	DECIMAL	0.00 <= x <= 100000.00
Fresh weight	Cassava_Leaves_Fresh_ g		1	5				Plot	Cas_FreW	DECIMAL	0.00 <= x <= 100000.00
Dry weight	Cassava_Leaves_Dry_w g		1	5				Plot	Cas_DryW	DECIMAL	0.00 <= x <= 100000.00
Air temperature, maximum	Air_temperature_max_ t °C		500	1	Frequency	Daily		Plot	AirTemp_ma	DECIMAL	-40.00 <= x <= 130.00
Air temperature, minimum	Air_temperature_min_ d °C		500	1	Frequency	Daily		Plot	AirTemp_mir	DECIMAL	-40.00 <= x <= 130.00
Solar radiation	Solar_radiation_MJ/(m² MJ/(m2.d)		500	1	Frequency	Daily		Plot	SolRad	DECIMAL	0.00 <= x <= 40.00
Precipitation amount, daily	Precipitation_amount_ c mm		500	1	Frequency	Daily		Plot	Precip_D	DECIMAL	0.00 <= x <= 10000.00
Wind speed, average daily	Wind_speed_average_ d km/h		500	1	Frequency	Daily		Plot	WindSpe_av	DECIMAL	0.00 <= x <= 35.00
Relative humidity, average	Relative_humidity_aver %		500	1	Frequency	Daily		Plot	RelHum_ave	DECIMAL	0.00 <= x <= 100.00
Evapotranspiration	Evapotranspiration_am mm		500	1	Frequency	Daily		Plot	EvaTrans_D	DECIMAL	0.00 <= x <= 10000.00
Soil moisture	Cassava_Soil_moisture_ %		3	10	Other	planting inte 0-20 20-50		Plot	Soil_Mois	DECIMAL	0.00 <= x <= 100.00
Organic carbon content in soil	Cassava_Organic_carbon %		3	10	Other	planting inte 0-20 20-50		Plot	Soil_OrgC	DECIMAL	0.00 <= x <= 100.00
Total nitrogen content in soil	Cassava_Total_nitrogen %		3	10	Other	planting inte 0-20 20-50		Plot	Soil_TotalN	DECIMAL	0.00 <= x <= 100.00
Available phosphorus content in soil	Cassava_Available_phosphorus %		3	10	Other	planting inte 0-20 20-50		Plot	Soil_Availabl	DECIMAL	0.00 <= x <= 100.00
Cation exchange capacity	Cassava_Cation_exchange_capacity Meq/100g		3	10	Other	planting inte 0-20 20-50		Plot	Soil_CEC	DECIMAL	0.00 <= x <= 10000.00
Soil pH	Cassava_Soil_pH_pH pH		3	10	Other	planting inte 0-20 20-50		Plot	Soil_pH	DECIMAL	0.00 <= x <= 14.00
Transplanting seedling density	Cassava_Transplanting_plant/m2		1	1				Plot	Tra_SeedDer	DECIMAL	0.00 <= x <= 100000.00
Seeding notes	Cassava_Seeding_notes		1	1				Plot	Seed_Notes	TEXT	TEXT
Residue management technique	Cassava_Residue_management_technique		1	1				Plot	Res_Tech	CATEGORICAL	Burning Incorporation S
Tillage implement type	Cassava_Tillage_implement_type		1	1				Plot	Til_Tool	CATEGORICAL	Chisel plough Disc plough
Tillage depth	Cassava_Tillage_depth_cm		1	1				Plot	Til_Depth	DECIMAL	0.00 <= x <= 100.00
Weeding technique	Cassava>Weeding_technique		1	1				Plot	Wed_Tech	CATEGORICAL	Chemical control Grazing
Weeding number of passes	Cassava>Weeding_number_of_passes		1	1				Plot	Wed_NumPa	INTEGER	0.00 <= x <= 100.00
Pest and disease control method	Cassava_Pest_and_disease_control_technique		1	1				Plot	PestC_Tech	CATEGORICAL	Biological control Chemical

Crop\_measurements

Crop\_management

Notes\_Deviations

Weather

Soil

TraitList





View collaborators

Can view and download

## Forms

Sort by: Alphabetical

Show inactive (0)

	Assign Field, Trial and/or Plot	255	Form	Jul 08, 2020	last 2 hours ago	▼
	Collect Soil Sample	1	Form	Jul 17, 2020	last 5 days ago	▼
	Describe Field	0	Webform	Jul 17, 2020	last 5 months ago	▼
	Measure Cassava Plot	0	Webform	Jul 08, 2020	last 8 months ago	▼
	Measure Potato Plot	132	Webform	Jul 17, 2020	last 2 months ago	▼
	Process Plant Samples	1	Webform	Jul 23, 2020	last a month ago	▼



# What is an ontology??

Formal naming and knowledge representation system for a domain

Concepts and categories of “things” defined and organized to show their hierarchies and relationships

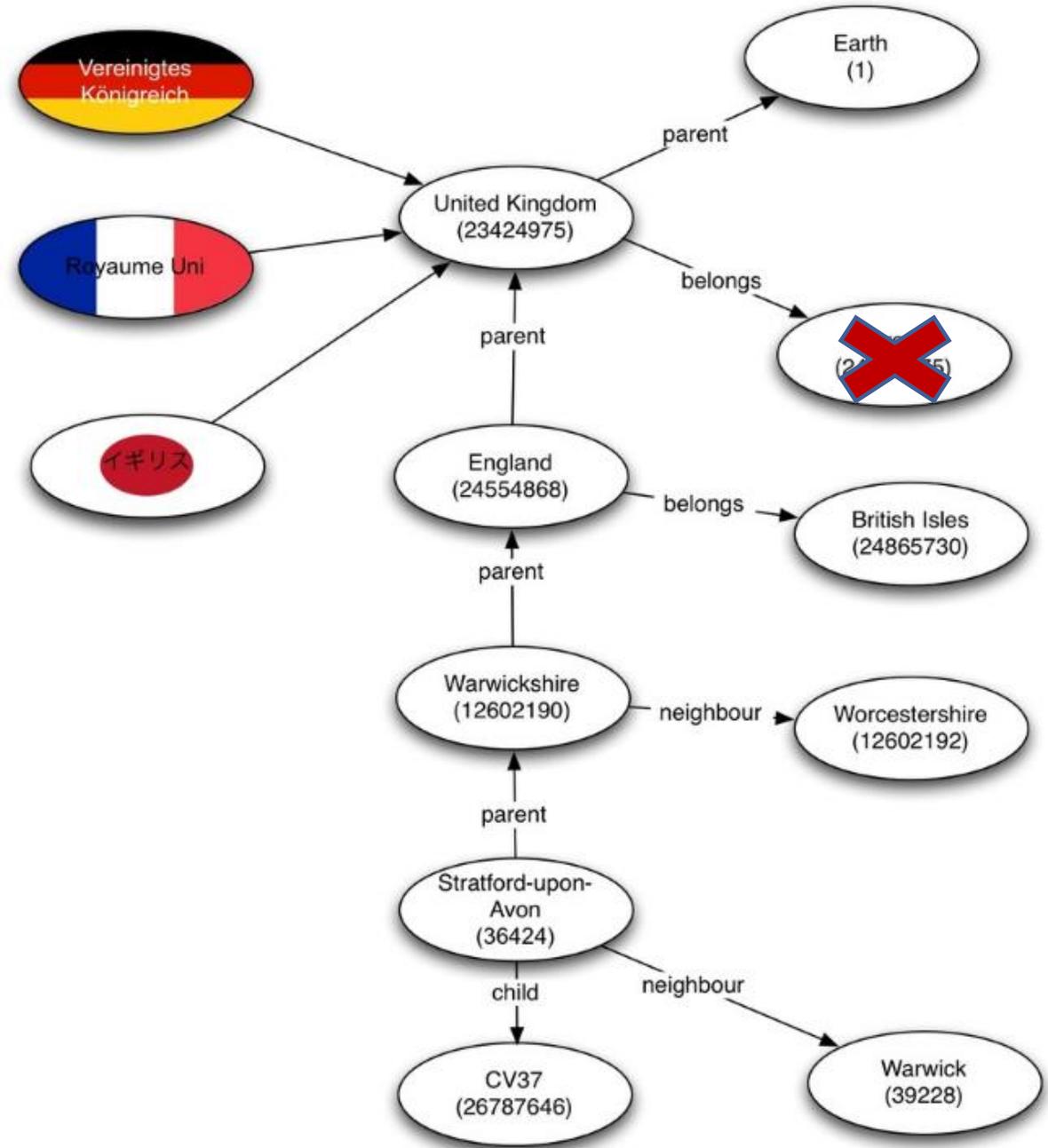
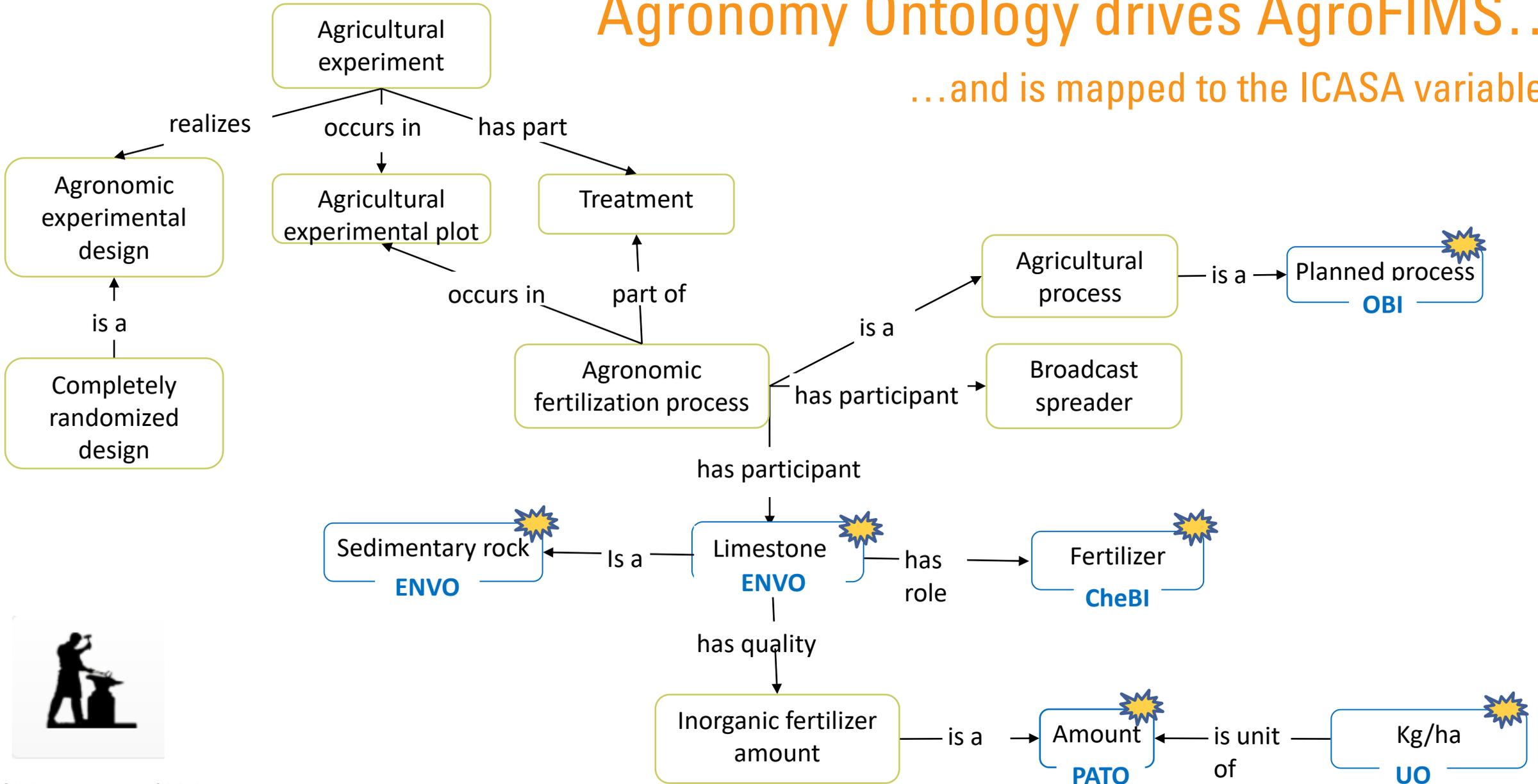


Fig. 3: Yahoo's WOID place ontology example for Stratford-upon-Avon. Source: 'Know Your Place: Adding Geographic Intelligence To Your Content', Gary Gale, AGI GeoCommunity 2009

# Agronomy Ontology drives AgroFIMS...

...and is mapped to the ICASA variables



# Quick Summary

The following summarizes each principle. See individual pages for details.

P1) **Open** - The ontology MUST be openly available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed in altered form under the original name or with the same identifiers.

P2) **Common Format** - The ontology is made available in a common formal language in an accepted concrete syntax.

P3) **URI/Identifier Space** - Each ontology MUST have a unique IRI in the form of an OBO Foundry permanent URL (PURL).

P4) **Versioning** - The ontology provider has documented procedures for versioning the ontology, and different versions of ontology are marked, stored, and officially released.

P5) **Scope** - The scope of an ontology is the extent of the domain or subject matter it intends to cover. The ontology must have a clearly specified scope and content that adheres to that scope.

P6) **Textual Definitions** - The ontology has textual definitions for the majority of its classes and for top level terms in particular.

P7) **Relations** - Relations should be reused from the Relations Ontology (RO).

P8) **Documentation** - The owners of the ontology should strive to provide as much documentation as possible.

P9) **Documented Plurality of Users** - The ontology developers should document that the ontology is used by multiple independent people or organizations.

P10) **Commitment To Collaboration** - OBO Foundry ontology development, in common with many other standards-oriented scientific activities, should be carried out in a collaborative fashion.

P11) **Locus of Authority** - There should be a person who is responsible for communications between the community and the ontology developers, for communicating with the Foundry on all Foundry-related matters, for mediating discussions involving maintenance in the light of scientific advance, and for ensuring that all user feedback is addressed.

P12) **Naming Conventions** - The names (primary labels) for elements (classes, properties, etc.) in an ontology must be intelligible to scientists and amenable to natural language processing. Primary labels should be unique among OBO Library ontologies.

P16) **Maintenance** - The ontology needs to reflect changes in scientific consensus to remain accurate over time.

## OBO Foundry Principles

High quality

Community-verified

Interoperable ontologies

Logically well-formed

Scientifically accurate



# AgrO content

<https://github.com/AgriculturalSemantics/agro>

- experiment
  - ▼ ● agricultural experiment
    - crop rotation experiment
    - crop yield experiment
    - ▶ ≡ inter-cropping experiment
    - long-term experiment
    - ▶ ≡ mono-cropping experiment
    - sequence experiment
    - ▶ ● short-term experiment

- object
  - agricultural implement
    - chemical pest control implement
    - ▶ ≡ fertilization implement
    - ▶ ≡ harvest implement
      - baler
      - binder
      - combine harvester
      - flail
      - ▶ ≡ mowing implement
        - scythe
        - sickle
        - tedder
    - ▶ ≡ irrigation equipment
    - ▶ ≡ land levelling implement
    - ▶ ≡ planting implement
      - dibbling stick
      - double disk opener planter
      - earth auger
      - hand dibbler
      - naven dibbler
      - rotary dibbler
      - single disc opener planter
      - ▶ ≡ sowing implement
        - broadcast spreader
        - drum seeder
        - furrow opener
        - hoe
        - ▼ ● seed drill
          - manual oilseed drill
        - tyne opener planter
    - ▶ ≡ residue management implement
    - ▶ ● tractor

## ● agronomic fertilizer

- ▼ ≡ inorganic fertilizer
  - calcium carbonate
  - ▶ ● limestone
  - magnesium carbonate
  - ▶ ≡ nitrogen fertilizer
  - NPK fertilizer
  - ▶ ≡ phosphorus fertilizer
  - ▶ ≡ potassium fertilizer
    - potassium chloride
    - potassium nitrate
    - sodium selenate
- ▼ ≡ organic fertilizer
  - alfalfa meal
  - ▶ ● animal manure (AGRO:00000079)
  - bagasse
  - biofertilizer
  - blood meal
  - bone meal
  - bovine dairy liquid manure
  - ▶ ● charcoal
  - ▶ ≡ compost
  - ▶ ● fish fertilizer
  - green manure
  - oil cake
  - ▶ ● oil meal
  - ▶ ● peat soil
  - ▶ ● sludge

# How do I easily make use of this brave, new data landscape?

## My team

Somebody I collaborate with (in KE)

Team lead (in BE)

Collaborator (in RW)

## GitHub

My repo

Our repo

Their repo

Somebody's repo

```
File Edit View Run Kernel Git Tabs Settings Help
~/hazard_layers /
Name Last Modified
aridity 2 months ago
chirps_cv 14 days ago
country_data 2 months ago
dry_days_future 13 days ago
dry_days_hist 13 days ago
ecocrop 3 months ago
heat_stress_days_future 12 days ago
heat_stress_days_hist 11 days ago
heat_stress_flips 2 months ago
hi 2 months ago
max_cons_dry_days_future 12 days ago
max_cons_dry_days_hist 13 days ago
mockup_tests 2 months ago
ppt_driest_month 2 months ago
ppt_driest_quarter 2 months ago
thi 2 months ago
aridity_thornthwaite.tar.bz2 2 months ago
chirps_cv.tar.bz2 14 days ago
dry_days_future.tar.bz2 12 days ago
dry_days_hist.tar.bz2 8 days ago
heat_stress_days_future.tar.bz2 10 days ago
heat_stress_days_hist.tar.bz2 10 days ago
heat_stress_flips.tar.bz2 2 months ago
hi.zip 2 months ago
max_cons_dry_days_future.tar.bz2 11 days ago

jovyan@456ce56792e2: ~$ ERA_analyses.R
# JRV
ERA analyses for adaptation atlas
#load
#install.packages("rworldmap")
#install.packages("micceadds")
#install.packages("rworldxtra")
require(anaologues)
require(rgdal)
require(raster)
require(maptools); data(wrld_simpl)
require(data.table)
require(Misc)
require(tidyverse)
require(ggplot2)
require(metR)
require(rworldmap)
require(doSNOW)
require(micceadds)
require(parallel)
requireforeach)
#analysis version
vr <- 4
# Set number of cores for parallel processing ---
Cores=8
# Run full or streamlined analysis?
DoLite<-T
#set directories - MAKE SURE YOU START CONSOLE FROM ANALOGUES FOLDER---
wd <- "~/work"
if(!dir.exists(wd)){dir.create(wd)}
cmdir <- paste(wd,"/ERA_analogues_modelling",sep="")
if(!dir.exists(cmdir)){dir.create(cmdir)}
#soildir <- paste(wd,"/soilgrids_global/Skm",sep="")
#soildir <- paste(wd,"/soilgrids_global/Skm",sep="")
```

## CGLabs

## Data

CGIAR data from GARDIAN

e.g., CMIP, MapSPAM, SoilGrids

IITA DV  
IFPRI DV  
...

Link

Download

SFTP transfer

Other datasets via internet; Google Drive (e.g. CHIRPs etc.)

My own or my Center's local data

## Computing resource

Amazon web services (AWS)

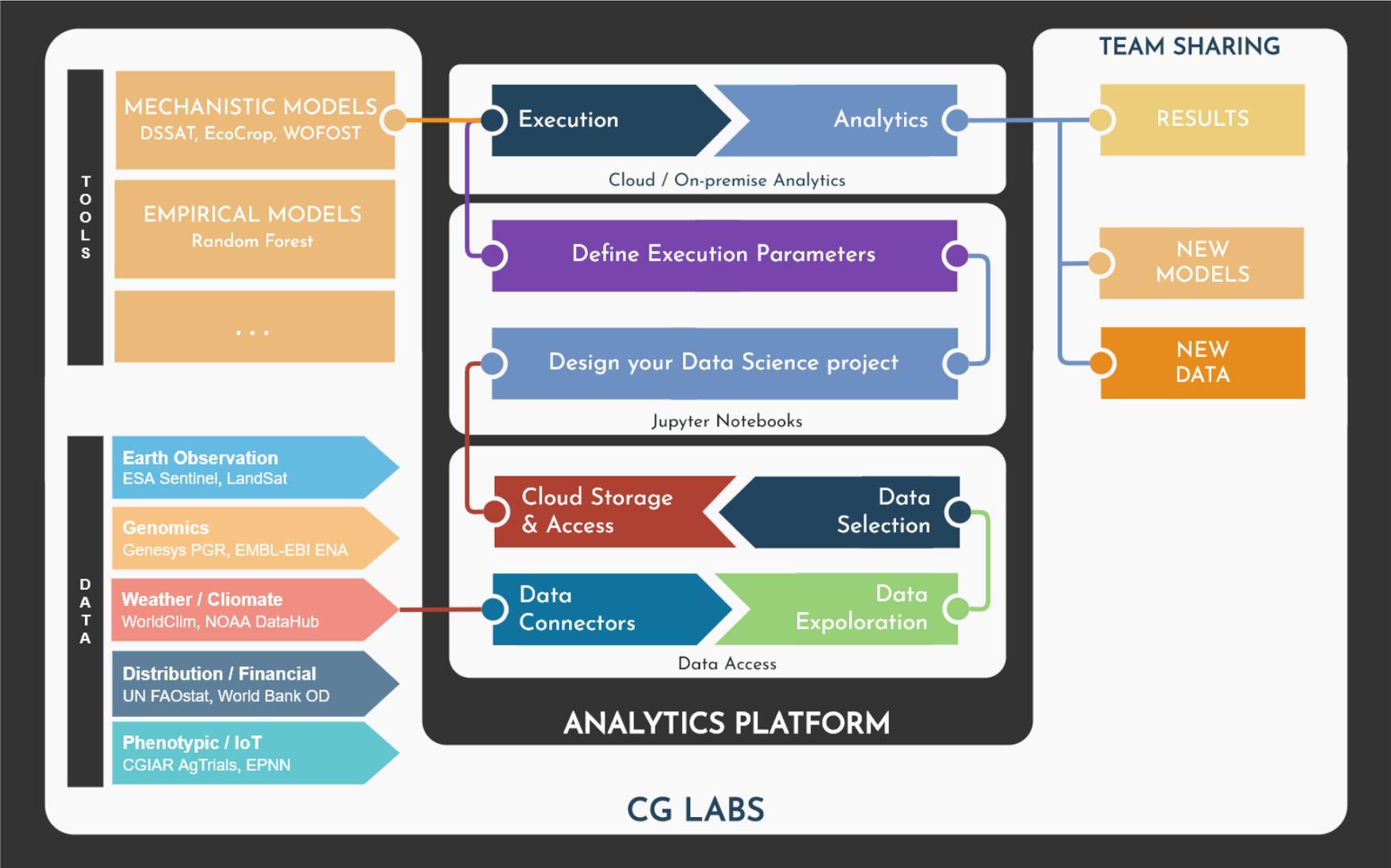
My center's computing resources

## Collaborative outputs?

spatial analyses, maps, simulations (DSSAT, WOFOST; QUEFTS, APSIM in 2021) ++

Modified from J. Ramirez Villegas

# CG Labs under the hood



# CG Labs in use: Crop suitability prediction

Download past and future climate data

Process 100s of models (ecocrop, random forest) with 40+crops and climate scenario combinations

```
# start all possible CMIP5 download at 2.5 min resolution
library(analogues)
data(cmip5_table)
source("~/work/0_getCMIP5.R")

# all possible options
vars <- c("tmin","tmax","tmean","prec","bio")
rcps <- c(2.6, 4.5, 6.0, 8.5)
years <- c(2030, 2050, 2070, 2080)
path <- datadir
gcm <- tolower(c("MOHC_HADGEM2_ES", "CESM1_CAM5", "GFDL_CM3", "MPI_ESM_LR", "MIROC_MIROC5"))

mlist <- cmip5_table$id[cmip5_table$model %in% gcm]

for (var in vars){
  for (rcp in rcps){
    for (year in years){
      lapply(mlist,
             function(m, var, rcp, year, path){
               .getCMIP5(var = var, rcp=rcp, model=m, year=year, res=2.5, path=path)},
             var, rcp, year, path)
    }
  }
}
```

download

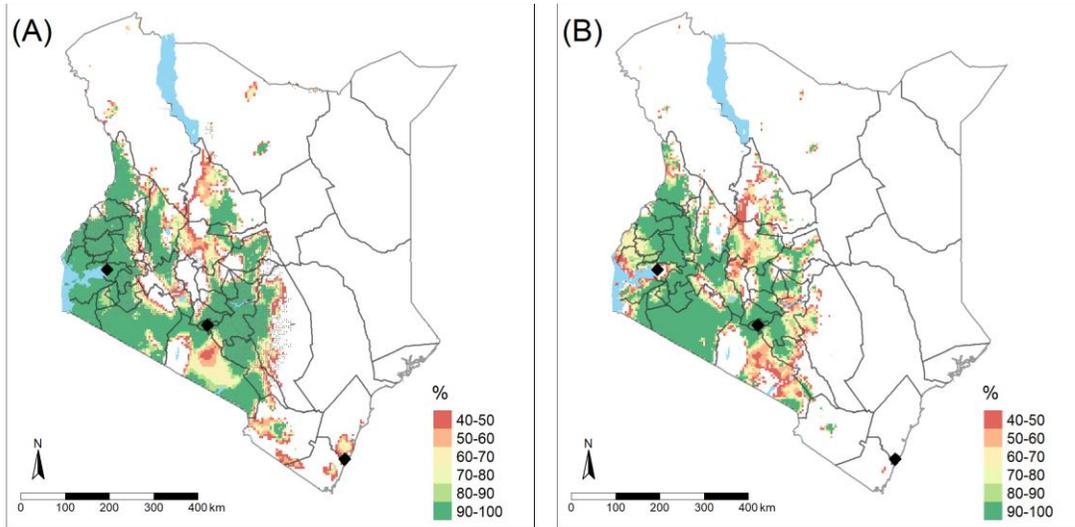
```
# run on worldclim
vars <- c("tmean","prec")
wc <- lapply(vars, function(var)getData('worldclim', var=var, res=10, path = "data/climate"))
ta <- rast(wc[[1]])
pr <- rast(wc[[2]])
```

ecocrop

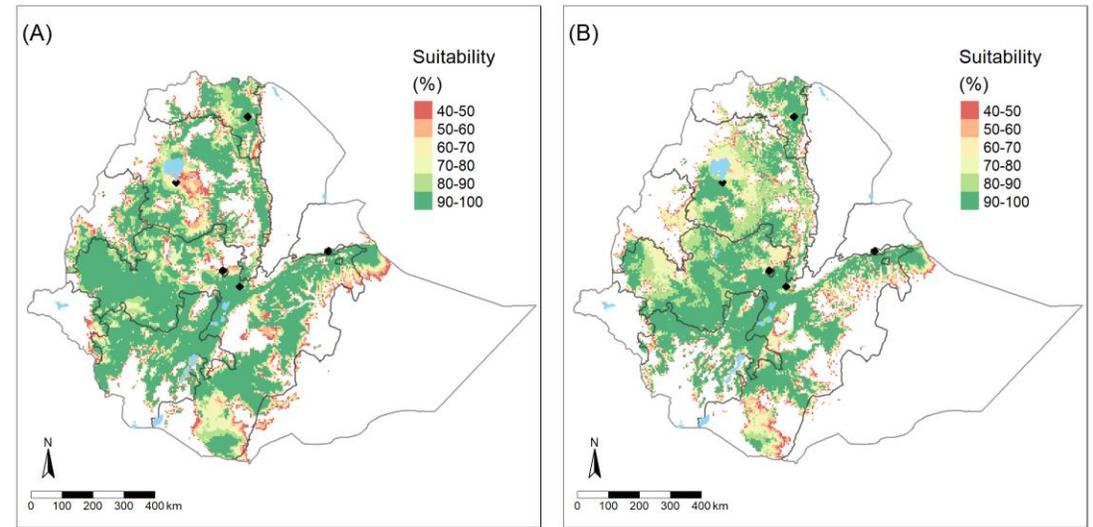
```
# based on dismo::ecocrop
runEcocropSingle(i,eco, ta, pr, "worldclim", outdir, res)
lapply(1:nrow(eco), runEcocropSingle, eco, ta, pr, "worldclim", outdir, res)
#parallel::mclapply(1:nrow(eco), runEcocropSingle, eco, ta, pr, "worldclim", outdir, res, mc.cores = 6, mc.preschedule = FALSE)
```

# CG Labs in use: Crop suitability prediction

Changes in common bean suitability (A) current; (B) Future



Kenya



Ethiopia

# CG Labs in use: Yield prediction at scale

## WOFOST and DSSAT simulation with large number of parameters

```
options(DSSAT.CSM="/home/jovyan/dssat-csm-os/build/bin")
dssatpkgs <- c("DSSAT", "Dasst", "lubridate", "stringr", "data.table", "raster")
lapply(dssatpkgs, require, character.only = TRUE)

# Load helper functions
basefuns <- list.files(file.path(indir, "funs"), pattern = ".R$", full.names = TRUE)
sapply(basefuns, source)

# GCM Models
mod <- c('BCC-CSM2-MR', 'EC-Earth3-Veg', 'GFDL-ESM4', 'IPSL-CM6A-LR', 'MRI-ESM2-0')

# dimension of domain
lon <- seq(from=29.125,to=40.125,by=0.05)
lat <- seq(from=-11.975, to=-0.975,by=0.05)
len <- length(lon)*length(lat)
comb <- data.frame(year = c(rep(2030,20),rep(2050,20)), mod = c(rep(mod,each=4),rep(mod,each=4)), sc = rep(c(126,245,370,585),10),
                  period=c(rep('202101-204012',20),rep('204101-206012',20)), co2=c(rep(c(440,446,454,456),5),rep(c(469,508,544,569),5)),
                  stringsAsFactors = FALSE)

for (i in 1:nrow(comb)){
  dfile <- readRDS(paste0('~/.work/dssat_input/',cntry,'/',comb$year[i], '/',comb$mod[i], '_ssp',comb$sc[i], '_r1i1p1f1_',comb$period[i], '.RDS'))
  # in parallel
  parallel::mclapply(1:nrow(dfile), runDSSATSingle, dfile,outdir,comb,i,indir, mc.preschedule = FALSE, mc.cores = 40)
}
```

```
library(Rwofost)
runWofostSpatial <- function(crop, datadir, iso, period, soil){
  cat("Processing ", crop, "\n")
  fdir <- file.path(datadir, paste0("weather_cell_", iso))
  ww <- list.files(path = fdir, pattern = glob2rx(paste0("cell_*", period, "*.rds")), full.names = TRUE)

  yp <- parallel::mclapply(ww, estPotential, crop, period, soil, mc.preschedule = FALSE, mc.cores = 10)
  yp <- data.frame(data.table::rbindlist(yp))

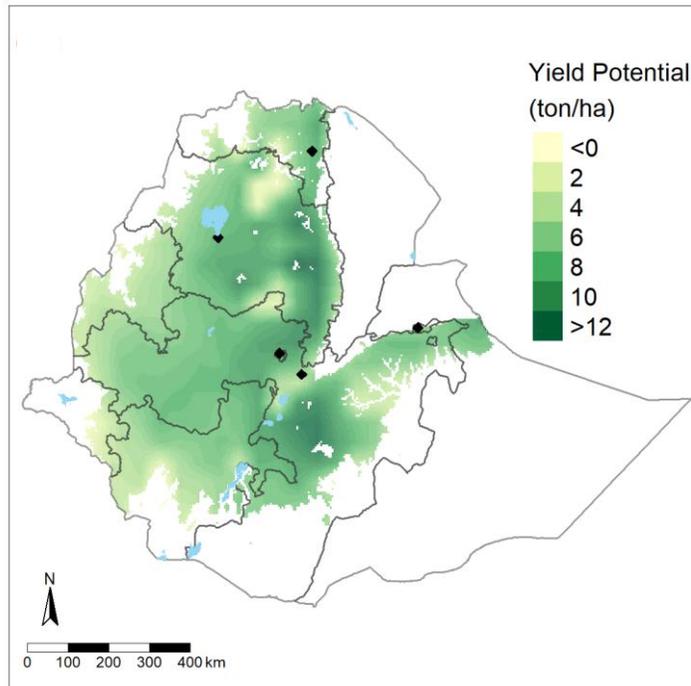
  outdir <- file.path(datadir, paste0("yield_potential_", iso))
  dir.create(outdir, FALSE, TRUE)
  fn <- file.path(outdir, paste0("yp_", crop, "_", soil, "_", period, ".rds"))
}
```

wofost

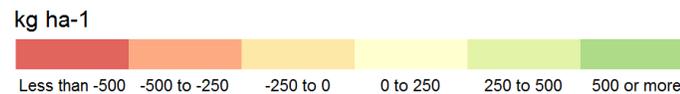
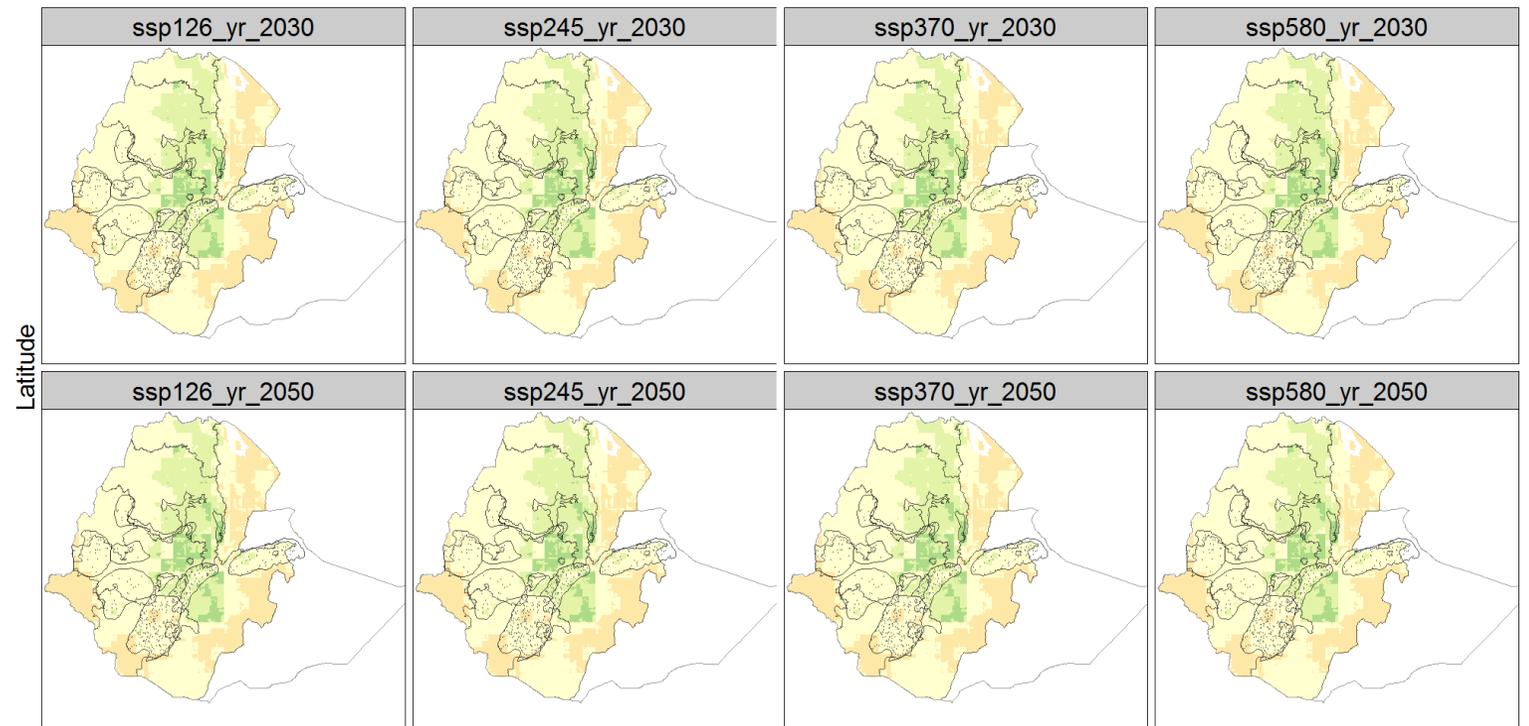
DSSAT

# CG Labs in use: Yield prediction at scale

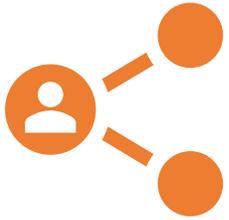
Wheat yield potential (WOFOST)



Effect of varying planting dates on bean yield (DSSAT)



# CG Labs at a glance



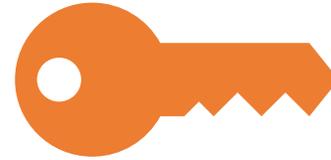
## Data sharing & discovery

Organization-wide  
secure data sharing,  
discovery services



## Collaborative analysis

Multi-faceted  
collaboration spaces at  
customizable levels  
(project teams, labs...)



## Security Management

Role-based access  
and Single Sign-On



## Private Data

Search and index data to  
enable secure web-based  
and API access to data  
assets

# Advantages of CG Labs?



## Scientist Empowerment

Ready-to-use datasets & tools allow your team to focus on analysis – and to collaborate



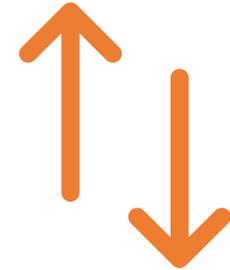
## Secure Access

Baked-in security protects your organization's data



## Seamless Integration

Flexible APIs simplify integration of toolsets (e.g. Weather Module uses NASA Power API)



## Flexible Deployment

Choose from on-premise, cloud or hybrid deployment

<https://gardian.bigdata.cgiar.org/labs.php>



About CG Labs

Frustrated at having your analytical workstream spread across multiple platforms?

Using Slack or a similar tool to collaborate with your team? Email/Dropbox/FTP etc. to exchange data while worrying about security? GitHub to manage code? Jupyter or other analytics platform for data analysis?

Try Collaborative GARDIAN Labs – offering interlinked features to find and securely exchange data, collaborate, manage code, and analyze!

Please sign in, using [Globus](#)

Choose language: [English \(US\)](#) ▼

Powered by [SCiO](#)

- ### Account settings
- Profile
  - E-Mail Summaries
  - Notifications
  - Settings
  - Security
  - Modules**

### User modules

Enhance your profile with modules.



**Calendar**  
Adds a calendar for private or public events to your profile and main menu.

**Disable** **Configure**



**Files**  
Adds files module to your profile.

**Disable**



**Weather Data**  
Genomic Data

**Enable**



**GARDIAN Index**  
GARDIAN User

**Disable**



**Globus Files**  
User Globus File System

**Disable**



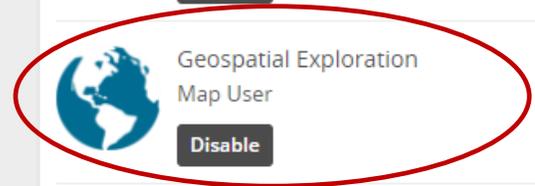
**Jupyter Lab**  
Jupyter User

**Disable**



**Geospatial Exploration**  
Map User

**Disable**





+  
-

Ethiopia

CNRM-CM6-1: max temperature

Download raw data :

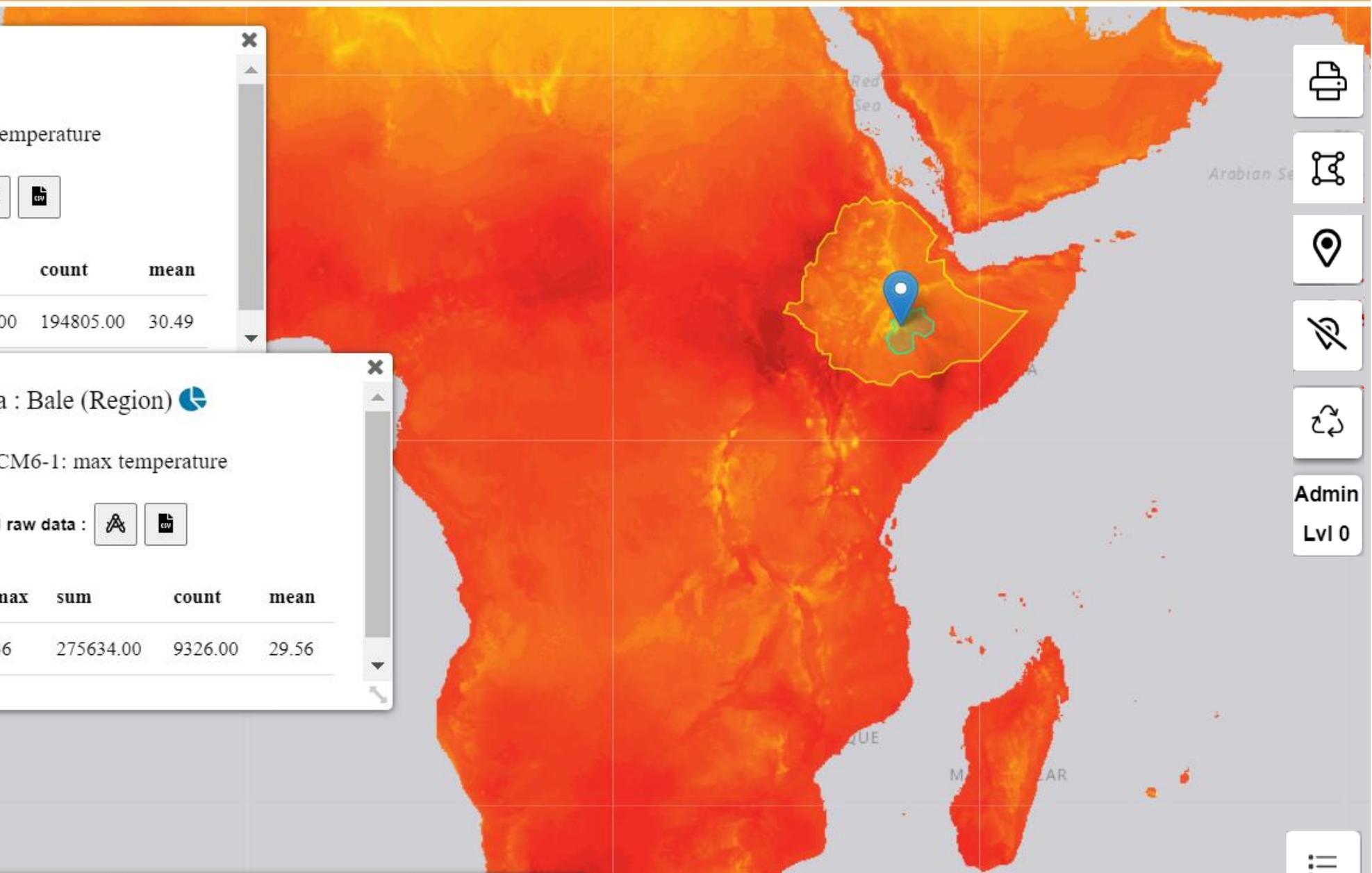
min	max	sum	count	mean
16	39	5940021.00	194805.00	30.49

Ethiopia : Bale (Region)

CNRM-CM6-1: max temperature

Download raw data :

min	max	sum	count	mean
16	36	275634.00	9326.00	29.56



SSP  
585

January 2030s (mean of 2020-2049)

1fps



# GARDIAN, CG Labs...: SCiO [<https://scio.systems>]



Pythagoras  
Karampiperis

Chief Executive Officer,  
founder



Panagiotis  
Zervas

Chief Operations Officer,  
partner



Sotiris  
Konstantinidis

Chief Technology Officer,  
founder



Antonis  
Koukourikos

Chief Innovation Officer,  
partner

---

# AgroFIMS: Developers, testers...

Ivan Perez (CIP)

Omar Benites (CIP)

Raul Eyzaguirre (CIP)

Pieter Pypers (IITA)

Meklit Chernet (IITA)

Jazmin Molano (CIP)

Elisa Salas (CIP)

Peter Craufurd (CIMMYT)

Balwinder Singh (CIMMYT)

Elizabeth Arnaud (Bioversity)

Henry Juarez (CIP)

**Celine Aubert (Alliance Bioversity, CIAT)**

Marie-Angélique Laporte (ABC)

Robert Hijmans (UC Davis)

Richard Ostler (Rothamsted)

Chris Villalobos (University of Florida)

Cheryl Porter (University of Florida)

Jeroen Huising (IITA)

Sylvain Delerce (CIAT/INRA)

Brian Lowe (Ontocale Inc)

Andrei Tudor (Ontocale Inc)

Vilma Hualla (CIP) ++



Platform for  
Big Data  
in Agriculture

[bigdata.cgiar.org](http://bigdata.cgiar.org)

