

FAIR Data – the Why and How

Medha Devare

Sr. Research Fellow (IFPRI), Big Data Module Lead



Platform for
Big Data
in Agriculture





The agriculture sector
needs to get smarter.

It needs to get digital.

How do we increase
global food quantity
and quality to feed
**nine billion people
by 2050?**

Improving access to
technology and
digitization in the
agricultural sector
can catalyse
sustainable
agricultural
systems.

How do we ensure smart, effective use of data?



CGIAR's Big Data Platform for Agriculture wants to harness Big Data, machine learning, and AI to turn **data into insight.**

Organize



Organize for FAIR
and open data,
analytics, and
capacity

Convene



Convene novel
partnerships and
shared learning

Inspire



Inspire with
examples and
innovations that
enhance impact



Why make data Findable, Accessible, Interoperable, Reusable (FAIR)?



Avoids unnecessary duplication in data collection

Results in cost saving



Enhances “data to insight” ability and options to address multi-faceted challenges with potential for increased impact

Facilitates core mission; RoI



Increases research visibility; demonstrates relevance via reuse

Enhances reputation and potential for more investment



Facilitates open science and accelerates innovation through ML, AI

Provides potential for data science-driven transformation of ag R4D



Meets requirements of donors, journals, socio-political environment (e.g. donor Open Data policies)

Complies with policies, G8, G20, country++ Open Data initiatives



Provides important resources for education and training

Democratizes access to knowledge; enhances capacity, self-sufficiency

What is the value of FAIR data?



(<http://gardian.bigdata.cgiar.org>)



ABOUT GARDIAN



GARDIAN

Discover agricultural data and publications across CGIAR and beyond

nutrition women



PUBLICATIONS

96175



DATASETS

3109

Search results:

Publications: **489**

Datasets: **59**

Pubmed Central: **62400**

ENA Accessions: **0**

GOV.UK: **3,623**



FILTERS

YEAR

All ▼

PROVIDER

All ▼

LOOK IN

All fields ▼

TYPE

All ▼

COUNTRY

All ▼

MAP

SORT BY RELEVANCE YEAR

2016

Scientific Publication

Gendered time-use patterns and effects on nutritional status of women and children in the semi-arid tropics: micro-level evidence from selected villages of India

ICRISAT

2005

Scientific Publication

Women, still the key to food and nutrition security

IFPRI

2018

Scientific Publication

Effect of nutrition education and psychosocial factors on child feeding practices: findings of a field experiment with biofortified foods and different women categories

CIP

2014

Scientific Publication

Smallholder farming and crop variety choice: Wheat variety choice in Pakistan

IFPRI

Search results:

Publications: 489

Datasets: 59

- | | |
|------|---|
| 2018 | Low birth weight in South Asia, trial dataset
<i>DFID</i> |
| 2015 | More pork by and for the poor: Catalyzing emerging smallholder pig value chains in Uganda for food security and poverty reduction
<i>ILRI</i> |
| 2018 | Dataset for: Feed the Future's Viable Sweetpotato Technology in Africa (VISTA) - Mozambique Nutrition Monitoring Survey 2017.
<i>CIP</i> |
| 2018 | Replication Data for: Promotion of Orange-Fleshed Sweet Potato Increased Vitamin A Intakes and Reduced the Odds of Low Retinol-Binding Protein among Postpartum Kenyan Women
<i>CIP</i> |
| 2018 | Replication Data for: MAMA SASHA Baseline Survey
<i>CIP</i> |
| 2016 | Replication data for: Gender roles and knowledge in plant species selection and domestication: case study in South and Southeast Sulawesi
<i>ICRAF</i> |
| 2018 | Rwanda Nutrition Survey, 2010
<i>IFPRI</i> |

00011000 1001 100
011011100 110 00/ 0
01010 00 10 011
011 011 11 011
000 100 01 010
1000000 00000100
00001101 0111001

ACCESS RIGHTS

License



Terms of use

[VIEW](#)

LINKS



[DOI](#)

DATASET

His and Hers, time and income: How intra-household dynamics impact nutrition in agricultural households

AVAILABLE FROM

International Center for Tropical Agriculture (CIAT)

AUTHORS

[Twyman Jennifer](#), [Useche Pilar](#), [González Carolina](#), [Talsma Elise](#), [Lopera Diana Carolina](#)

11001001101101110011110
00111001001101110001100
110011101001101001110
10111001100111010011011
11010011100111001101101
10001100110011001100
0110111001110001100110
010101100110011011011
01110111111111011
00110010101010
1011000100100100100
00001101011100110110011



[VIEW LARGE](#)

FAIR COMPLIANCE



F = 4.75 / 5

A = 3.76 / 5

I = 3.60 / 5

R = 4.04 / 5

[VIEW METRICS](#)

SUMMARY

Understand how dietary diversity is impacted by intra-household decision making processes related to income, nutrition information and time allocation (with primary focus on principle men and women in the household). In particular, we will examine how nutrition information, income, and time allocation impact food consumption.

Furthermore, we will analyze whether households where women have more decision-making power, as measured by choice experiments, systematically differ in their actual consumption patterns from households where men's preferences are more highly represented in decision-making.

We will use two types of data collection instruments:

- Surveys (Household and individual level)
- Hypothetical and real choice experiments

KEYWORDS

information carrier higher-level landform households
processes men nutrition time power

DATASET FILES



01. Household data collection Instrument.pdf

application/pdf



02. Individual data collection Instrument.pdf

application/pdf

RELEVANT PUBLICATIONS

- [What was the impact of dairy goats distributed by the Crop-Goat project in Tanzania?](#)
- [Poverty, household food security, and nutrition in rural Pakistan](#)
- [An integrated economic and social analysis to assess the impact of vegetable and fishpond technologies on poverty in rural Bangladesh](#)
- [Early childhood nutrition, schooling, and sibling inequality in a dynamic context, evidence](#)



How do I make my data FAIR?



Make your data:

Findable: persistent identifier; rich metadata; good documentation



	Variable	Description	Type	Unit	Average	Median	Minimum	Maximum	St Dev	No. missing values
SITE	Country name	Name of country in which experiments were conducted	Text	Text						
	SIMLESA Site name	Name of site eg district where experiments were	Text	Text						
	Farmercode	Farmner serial number	Numeric	Numeric						
	Treatment name/code	Refers to Cropping system used. CA= Conservation Agriculture; CP=conventional practice in the local country context	Text	Text						
	Legume association	Refers to legume intercropped or rotated with maize. INT=intercrop; SOLE= crop grown on its own; ROT=crops rotated	Text	Text						
	Geometry	How the field was prepared for planting. Options = flat, mounts, beds, ridges	Text	Text						
	Tillage Practice	Type of tillage system used. CP= conventional ploughing; R/F=conventional ridge and furrow; D/S=	Text	Text						
SITE DESCRIPTION	Season	Year of harvesting	Date	YYYY	2012.484	2012	2010	2016	1.74917354	1
	Slope	General estimate of slope of fields used for the trial	Numeric	%	3.200745	2.5	1	5.5	1.74076277	61
	Total rainfall	The total amount of rainfall received in that cropping season in mm	Numeric	mm	774.4452	729	213	1865.3	427.195465	178
	Rainfall in the first 30 days after planting	Rainfall in the first 30 days after planting	Numeric	mm	215.0086	236	26	483	115.550559	208
	BD 0-20 cm	Soil bulk density in g/cm ³	Numeric	g/cm3	1.388691	1.4	1.16	1.48	0.04504259	408
	Textural class	Refers to broad soil classification in terms of fineness or coarseness of its texture in top 20 cm eg clay loam, clay, silty loam, sand	Text	Text						
	Sand 0-20 cm	% soil content by weight constituting the sand fraction	Numeric	%	29.65865	27	17	55.5	10.3819011	298
	Silt 0-20 cm	% soil content by weight constituting the silt fraction	Numeric	%	28.26323	19	19	59.26553	14.6012951	298
Clay 0-20 cm	% soil content by weight constituting the clay fraction	Numeric	%	42.07643	54	12.5	54	16.632	298	
TREATMENT DESCRIPTION	Maize cultivar	Maize variety used	Text	Text						
	Legume species	Type of legume crop used	Text	Text						
	Legume cultivar	Legume crop variety name	Text	Text						
	Surface mulch type 1	Material used as residue cover	Text	Text						
	Quantity surface mulch 1	rate of surface residue cover application	Numeric	kg/ha	0.528034	0.75	0	1	0.41476683	168
	Surface mulch type 2	In cases where mixed residue cover types are used, this indicates the second type of mulch used	Text	Text						
	Quantity surface mulch 2	Rate of surface residue cover type 2 application	Numeric	kg/ha						
	Basal fertilizer type	Name of basal fertilizer used	Text	Text						
	Basal fertilizer quantity	Rate of basal fertilizer application	Numeric	kg/ha	59.92057	100	0	125	52.3231753	78
	DAP (P&N)	Phosphorus content of DAP basal fertilizer	Numeric	kg/ha	36.5838	46	0	69	24.2365779	226
		Nitrogen content of Diammonium phosphate fertilizer	Numeric	kg/ha						
	NPK	Phosphorus content of applied NPK basal fertilizer	Numeric	kg/ha						
		Nitrogen content of applied NPK fertilizer	Numeric	kg/ha						
Urea	Nitrogen content of applied Urea fertilizer	Numeric	kg/ha							

Make your data:

Findable: persistent identifier; rich metadata; good documentation

Accessible: unrestricted license; access to metadata and physical files

Interoperable: speak the same language across information resources
(use metadata, controlled vocabularies/ ontologies)

Reusable: composite of the other dimensions: $R = (F+A+I)/3$

FAIR guidelines available at: <https://guardian.bigdata.cgiar.org/metrics.php#!/>



HOME

SEARCH

DATA EXPLORATION

ABOUT

SERVICES

ANALYTICS

Search for



All

Open

Restricted

And

Or

SEARCH

ACCESS RIGHTS

License



Terms of use

[VIEW](#)

His and Hers, time and income: How intra-household dynamics impact nutrition in agricultural households

DATASET

AVAILABLE FROM
CIAT

FAIR COMPLIANCE



F = 4.51 / 5

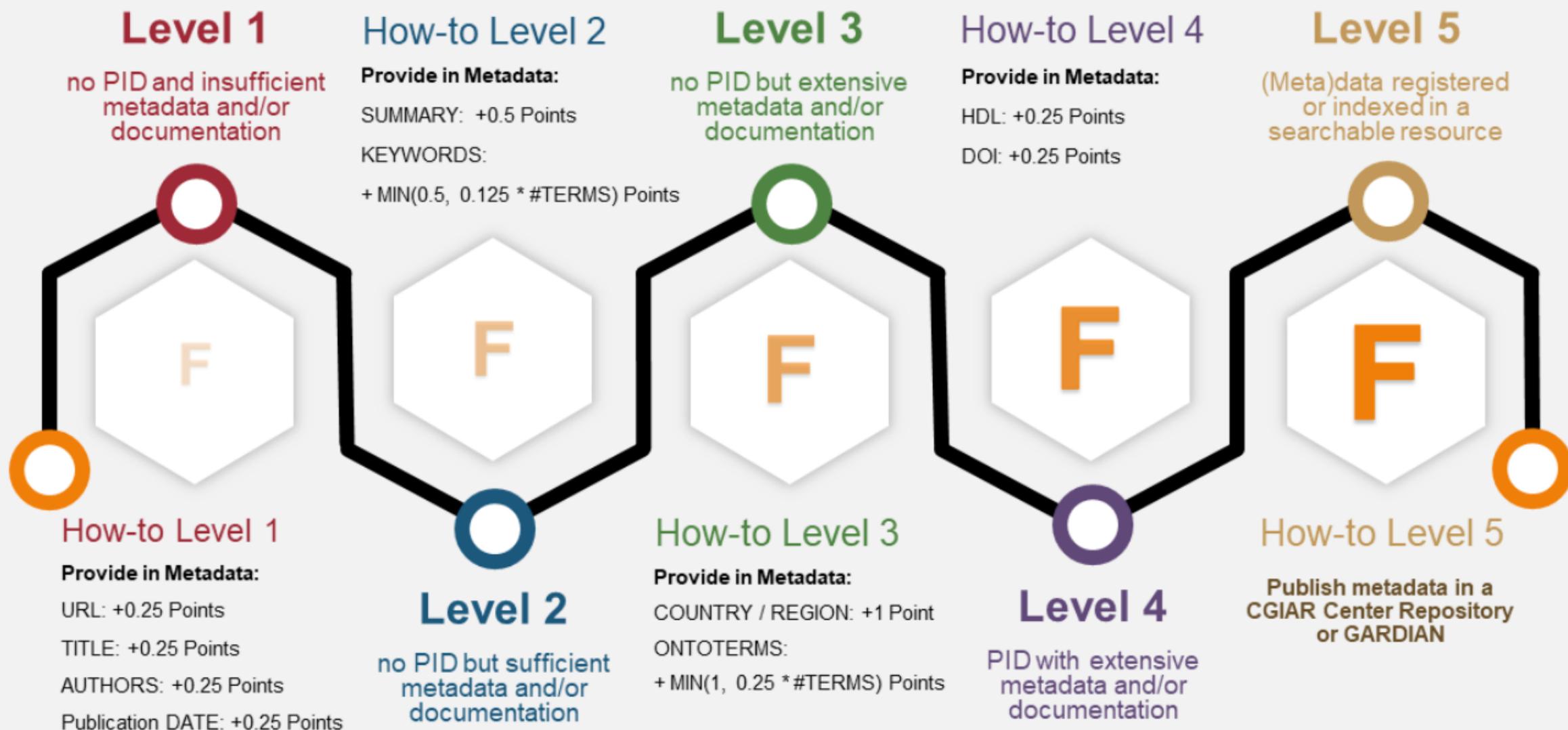
A = 4.70 / 5

I = 3.83 / 5

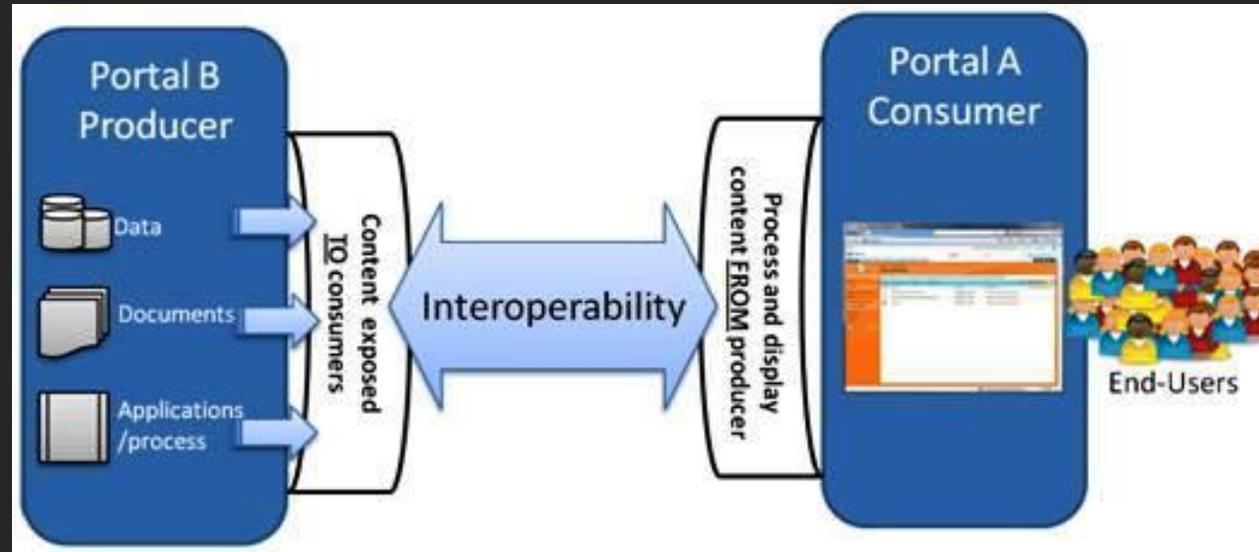
R = 4.35 / 5

[VIEW METRICS](#)

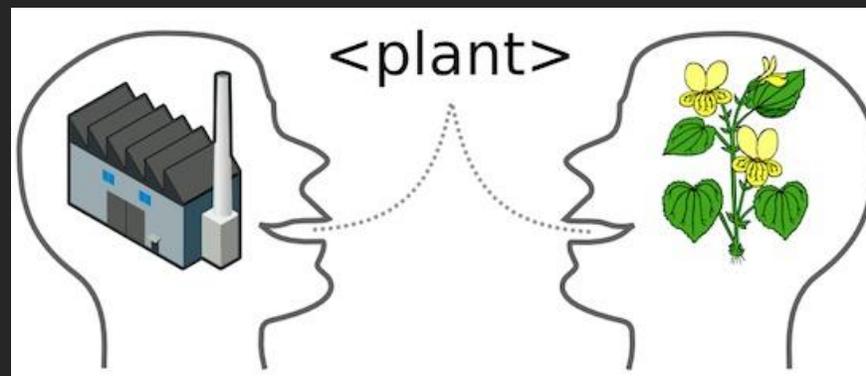
FINDABLE



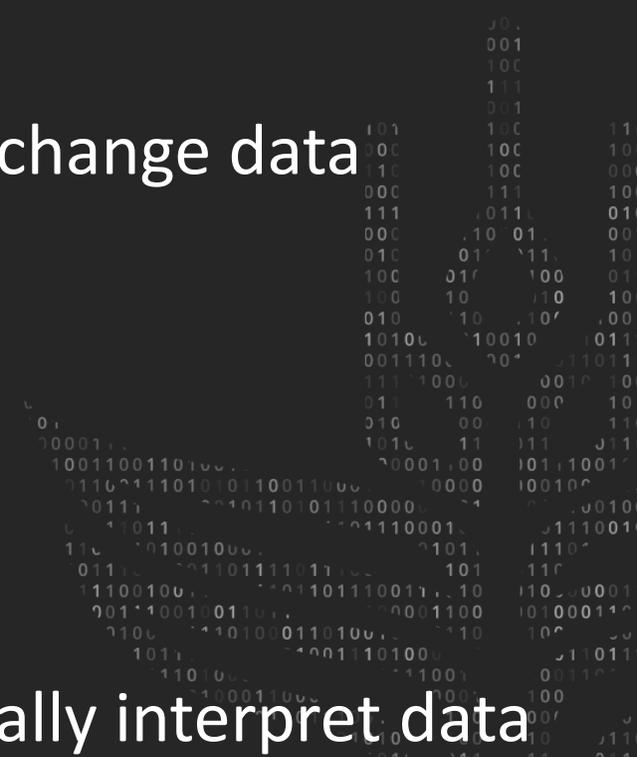
Interoperability



Syntactic interoperability: machines communicate and exchange data



Semantic interoperability: ascribe meaning to and automatically interpret data
(ontologies, common vocabularies, etc)



Interoperability

Traits, methods and scal

DOWNLOAD SHOW OBSOLETE TERMS

- Rice traits
 - Abiotic stress is_a
 - alkali injury is_a
 - cold tolerance is_a
 - drought injury is_a
 - drought recovery
 - heat injury is_a
 - iron toxicity injury
 - phosphorus deficiency
 - salt injury is_a
 - submergence tolerance
 - zinc deficiency injury
 - Agronomical is_a

Crop C

Home Ab

New: Tra

- The Crop
- the varia

To deve

- Download
 - the Trait
 - the 1st v
 - To get th
- For more

Search

Latest

Gene

FAO/Bio
terms
FAO/Biover
standard, d
M

Germplasm
germplasm

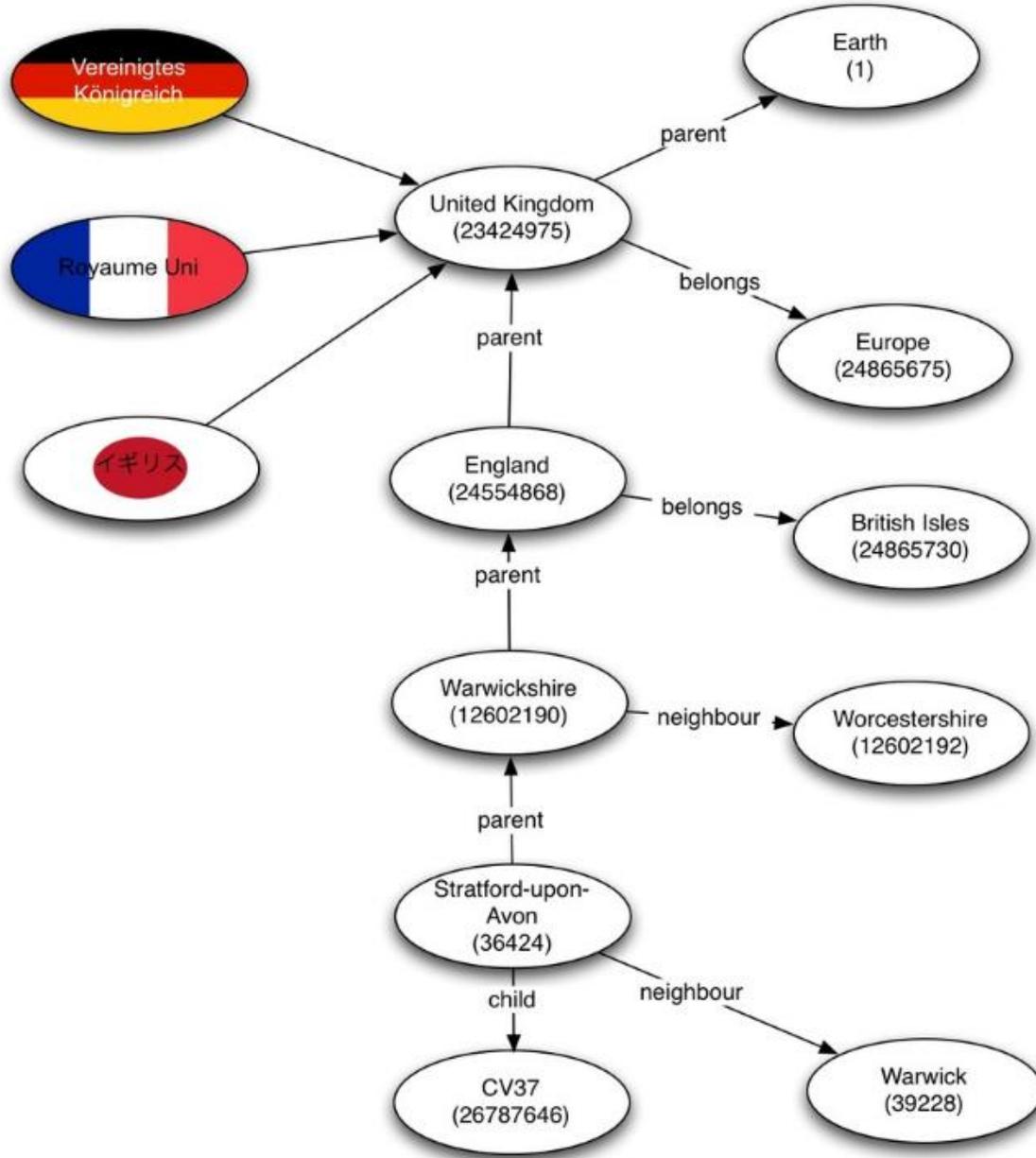
ICIS germ
ICIS gemp

Phen

Banana 3
Banana Tra
characteriz

Barley Tr
ICARDA - E

Barley Tr
Barley Trait
Plant Gene



ontology
agricultural data

to display

link ▾ General 0 Comments

ationtool.org

ter Login

Trait Dictionary

/ codes

ods

stages of
iantic
ession and
s.

stages of
iantic
ession and
s.

stages of
iantic
ession and
s.

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



diet therapy MeSH Qualifier Data 2019

Details Concepts

MeSH Qualifier	diet therapy
Unique ID	Q000178
Annotation	subhead only; for dietary & nutritional management of a disease by a physician; not for self diets; not for vitamin or mineral supplements prescribed by a physician (= /drug therapy); indexing policy: Manual 19.8.24; DF: /diet ther or /DH
Scope Note	Used with disease headings for dietary and nutritional management of the disease. The concept does not include vitamin or mineral supplements, for which "drug therapy" may be used.
Entry Version	DIET THER
Abbreviation	DH
Entry Term(s)	dietary management nutritional management
Online Note	search policy: Online Manual; use: main heading/DH or DH (SH) or SUBS APPLY DH
History Note	75; used with Category C & F3 1975 forward
Date Established	1975/01/01
Date of Entry	1974/10/23

Educação nutricional (pt), образование в области правильного питания (ru),
vzdelávanie v oblasti výživy (sk), Educación nutricional (es), การศึกษาด้านโภชนาการ

Interoperability – Linked Open Data

Legend

Cross Domain

Geography

Government

Life Sciences

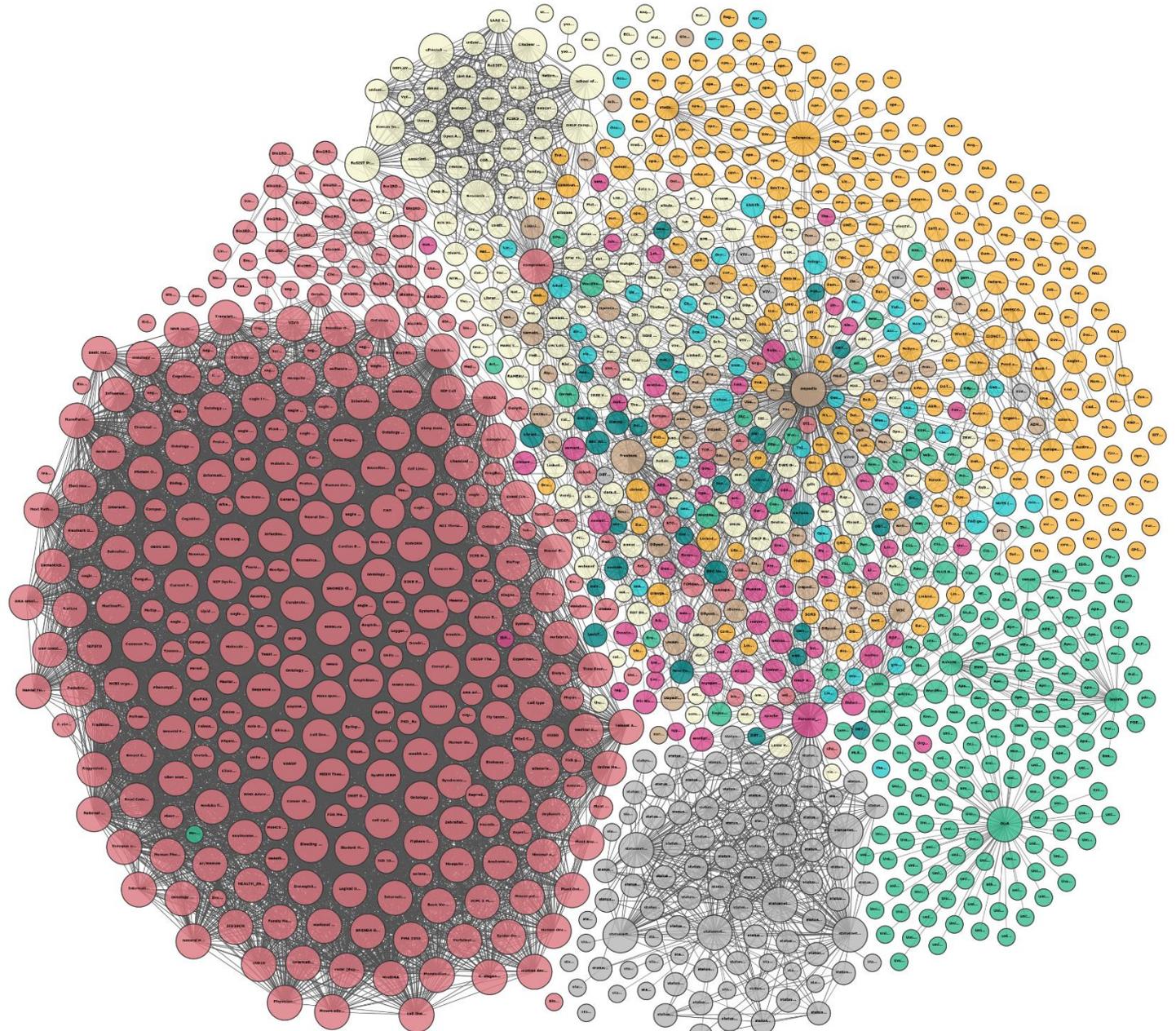
Linguistics

Media

Publications

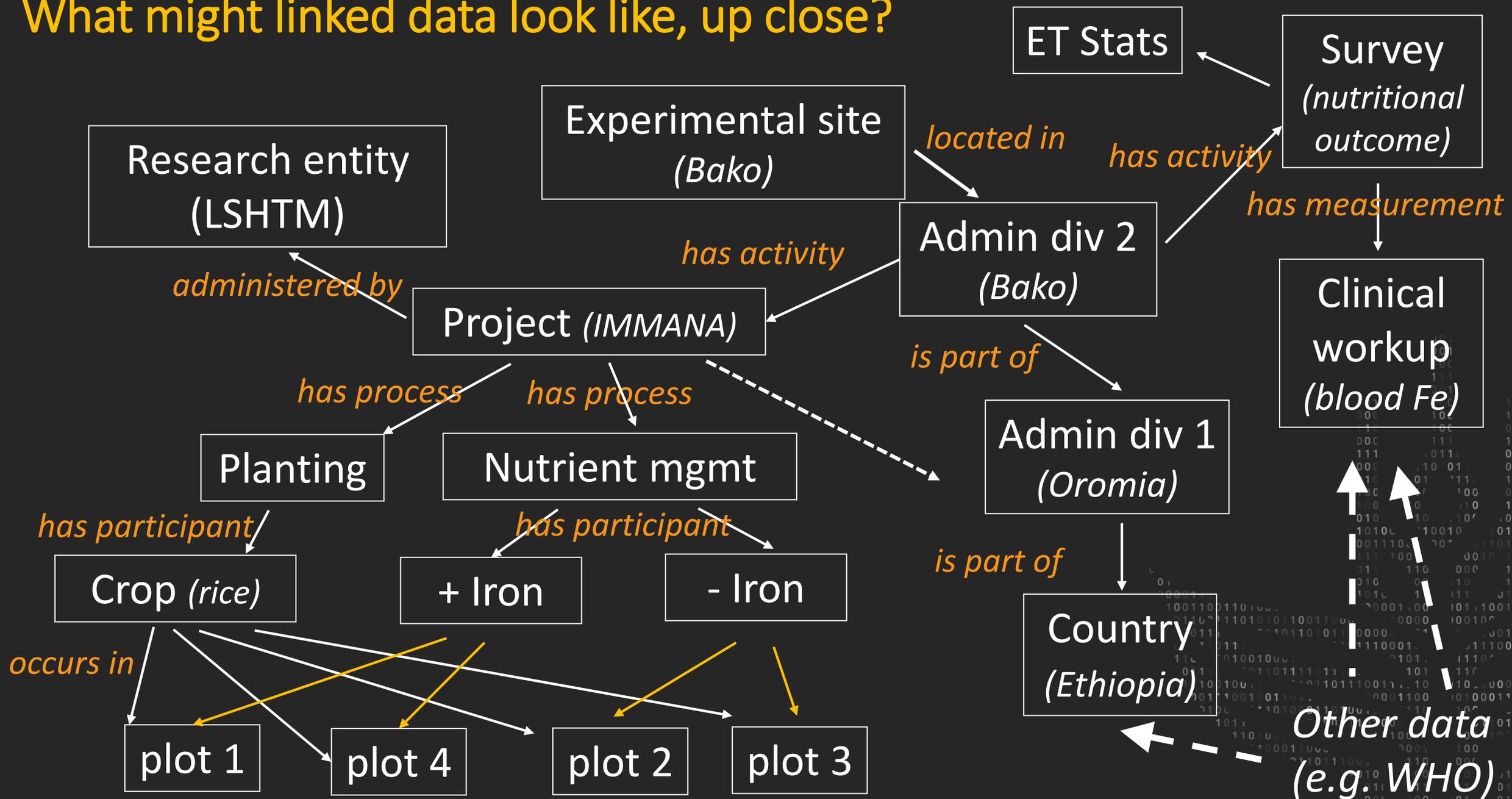
Social Networking

User Generated



Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentsch and Richard Cyganiak. <http://lod-cloud.net/>

What might linked data look like, up close?



Leveraging reusable data





Georeferenced Data



Semantically-enabled Data

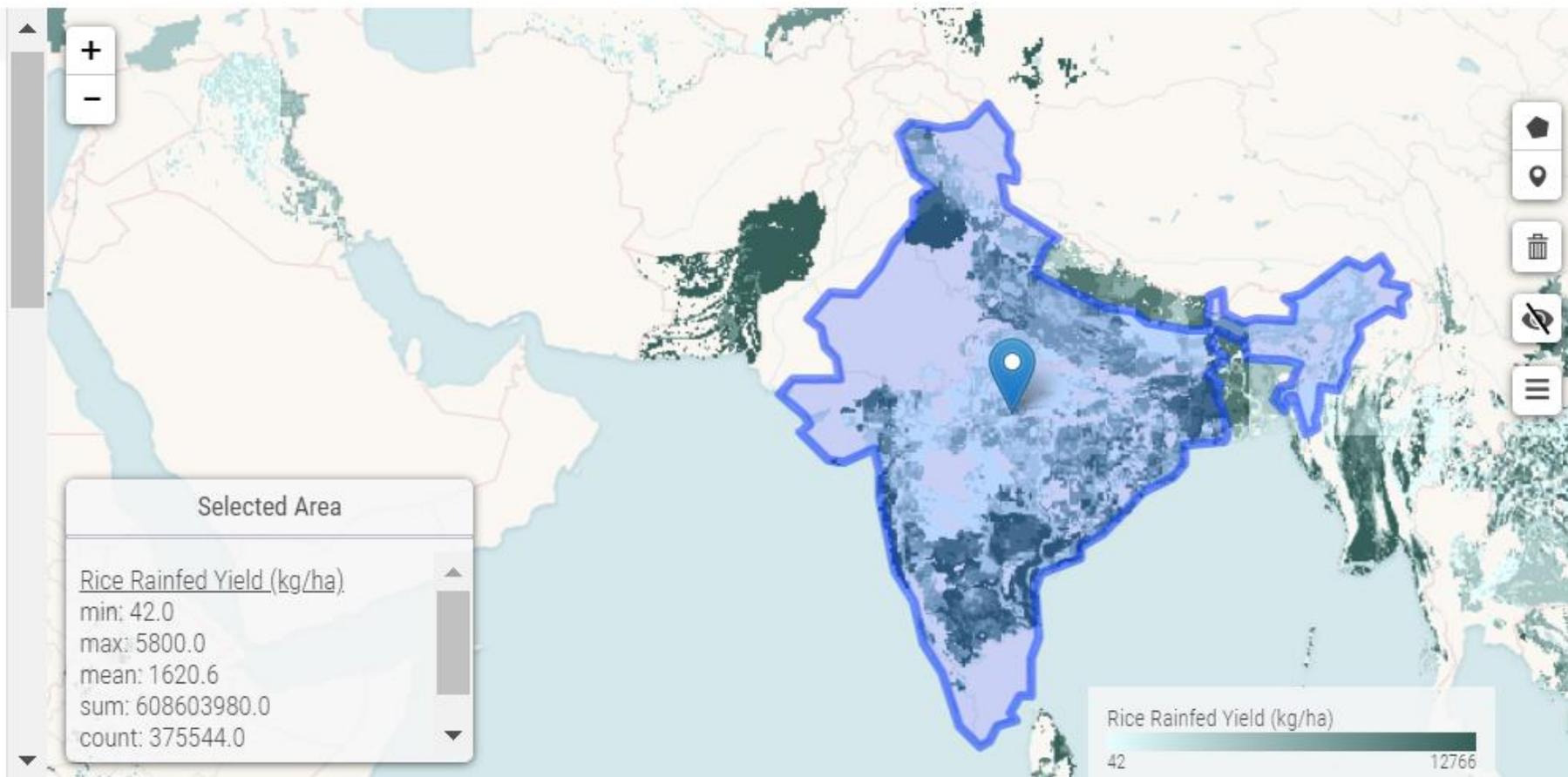
Explore crop production data for 2005, and collection sites for a total of about 4 million accessions from the Genesys database, stored in genebanks worldwide. Crop production data for 2010 and 2015 will also be available soon!

▼ Crop Production Stats 2005

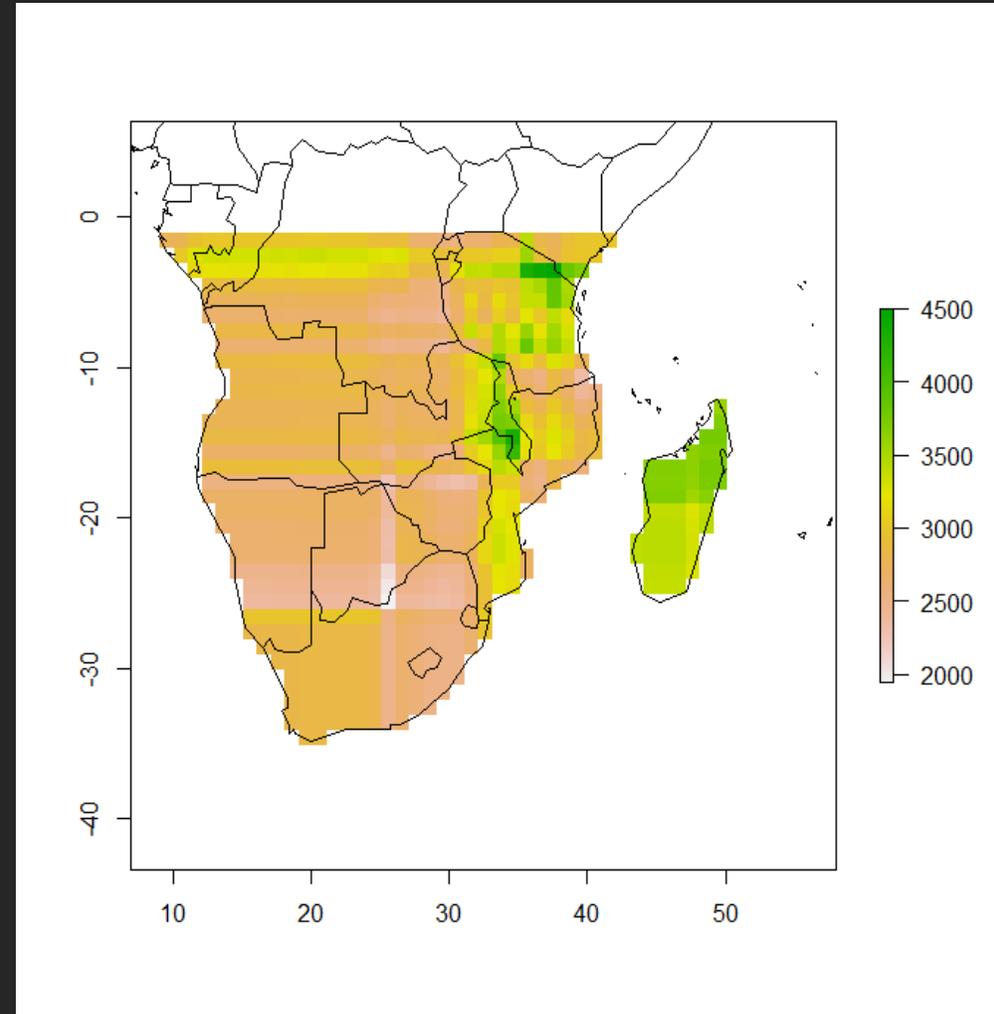
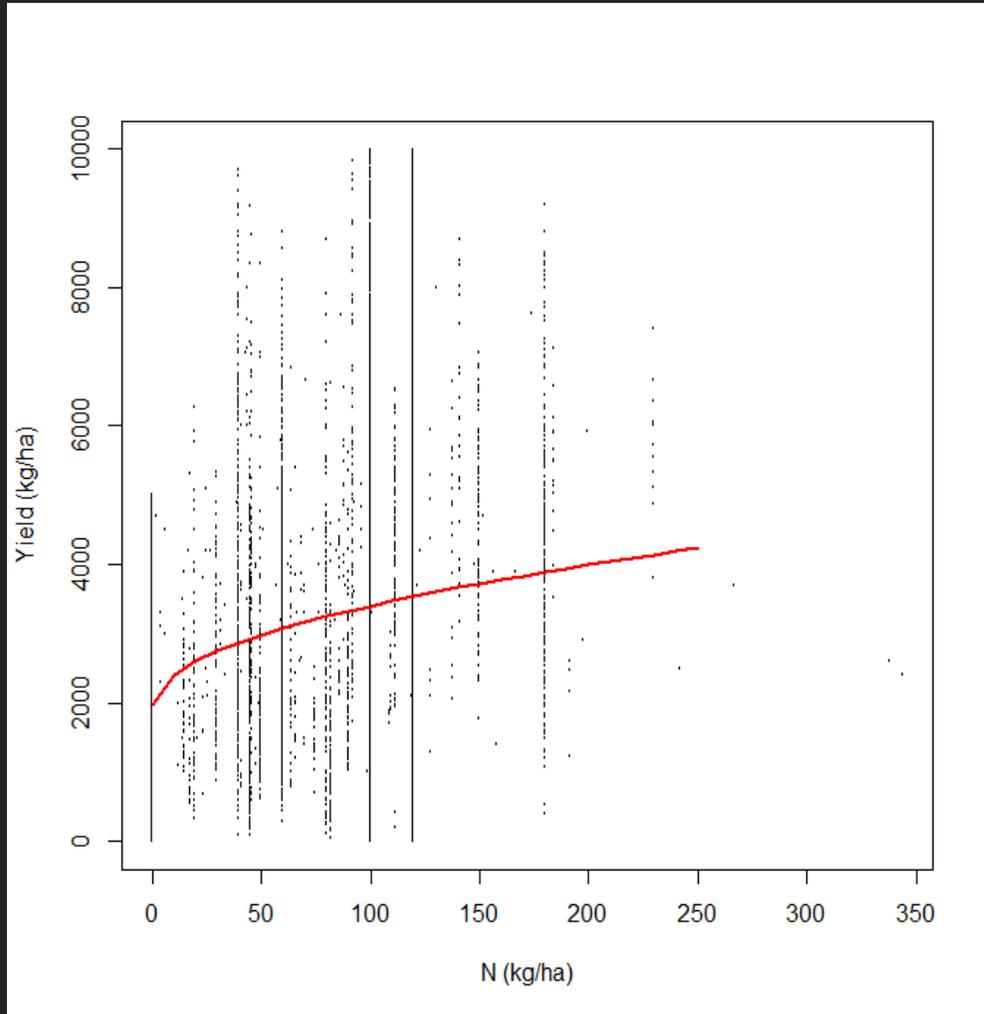
▼ Rice

- Irrigated Harvested Area (ha) ⚙
- Irrigated Production (mt) ⚙
- Irrigated Yield (kg/ha) ⚙
- Rainfed Harvested Area (ha) ⚙
- Rainfed Production (mt) ⚙
- Rainfed Yield (kg/ha) ⚙
- Total Harvested Area (ha) ⚙
- Total Production (mt) ⚙
- Total Yield (kg/ha) ⚙

▶ Maize



Using GARDIAN to understand variation in crop response to fertilizer across Sub-Saharan Africa



- Collect data digitally
- Add metadata use a standard schema: WHO? WHEN? WHAT? HOW? WHERE??
- Georeference data at collection – keeping privacy/ethics in mind
 - * i.e. “blur” data subject locations when they may be at risk



- Add abstract, rich summaries/descriptions
- Annotate resource with ontology terms
- Describe data variables with ontology terms
- Add an unrestricted use, machine-readable license to the resource
- Upload data to a public repository that allows itself to be searched/harvested



Questions?

Please contact Medha Devare (IFPRI Sr. Research Fellow and Module Lead) m.devare@cgiar.org

OR, come to the **Open Data Drop-In breakfast session** tomorrow
(Evolve Room)

Thank you!



Platform for
Big Data
in Agriculture

bigdata.cgiar.org

