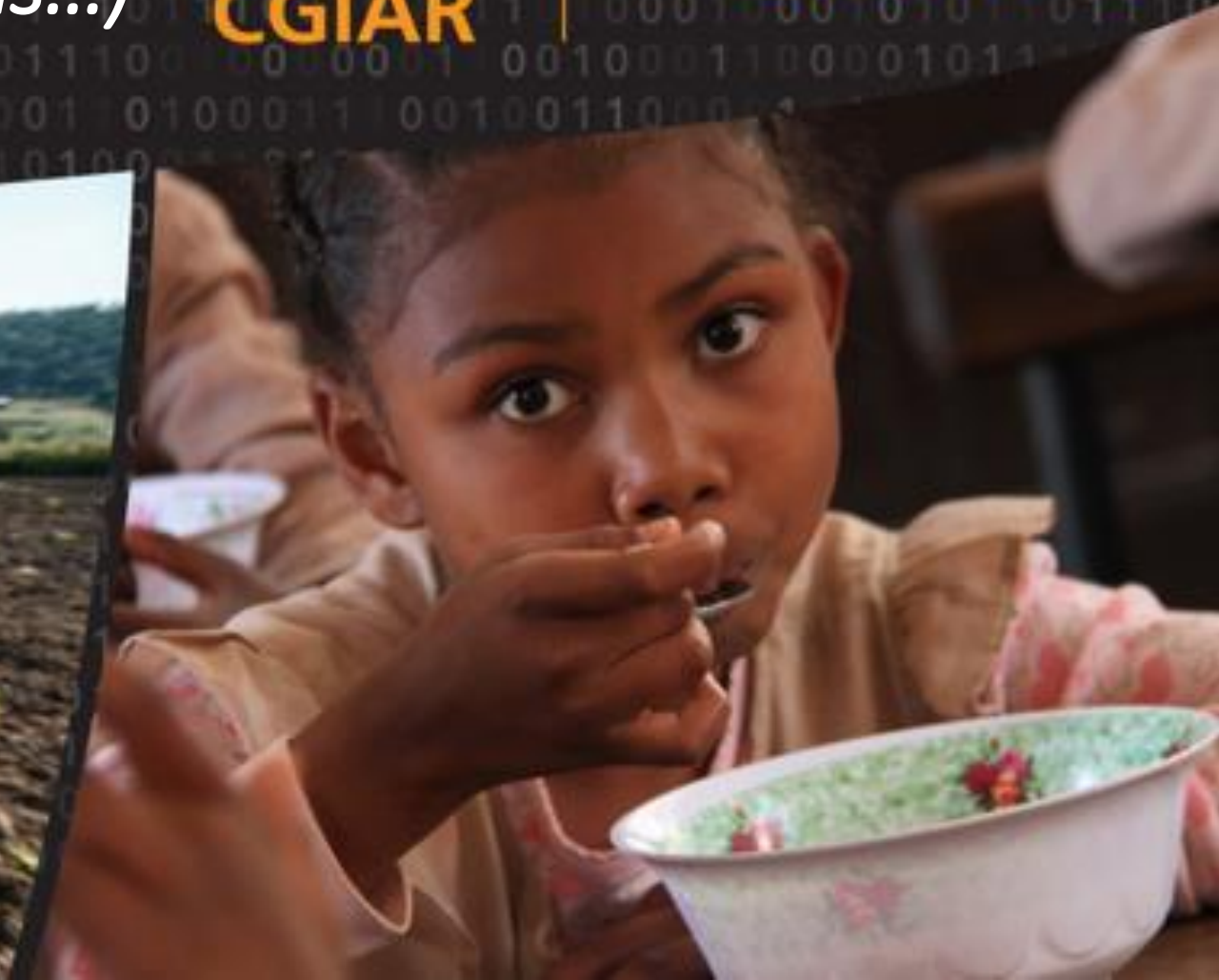
Standards Spur Agricultural Innovation

(Or: Your work impacts livelihoods…)

CGIAR | Platform for Big Data in Agriculture

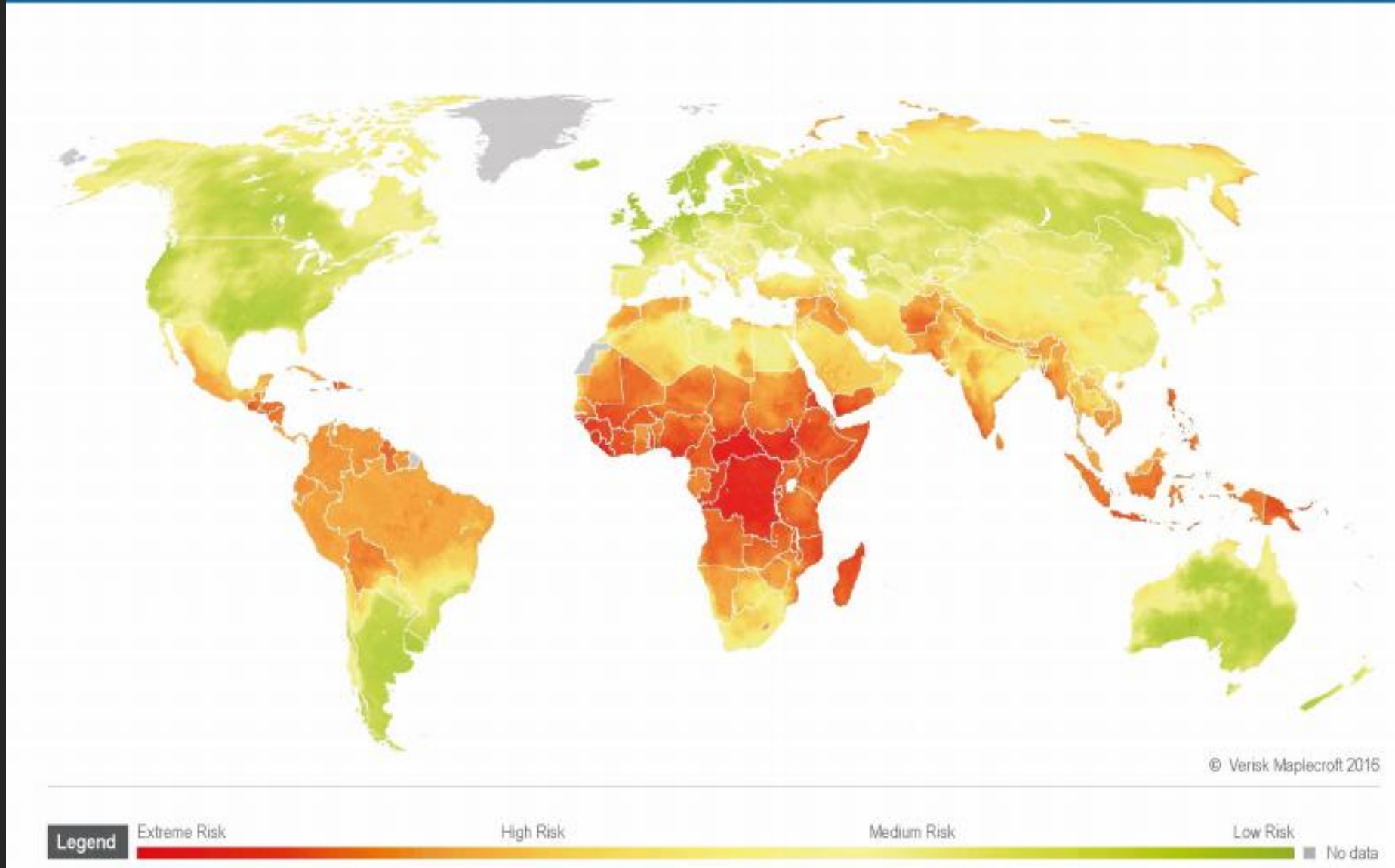
Medha Devare

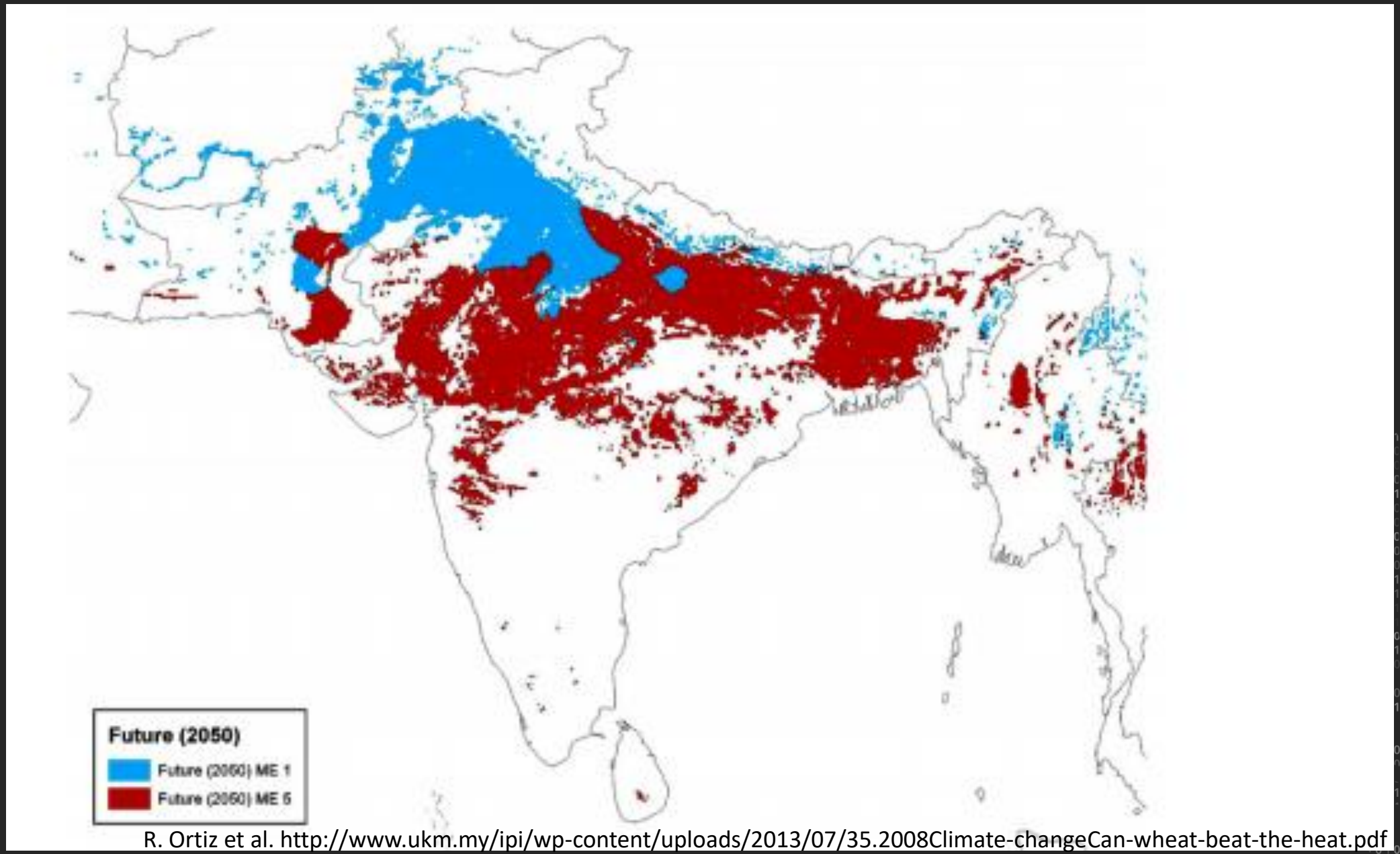
DCMI/TPDL, Sep 10, 2018



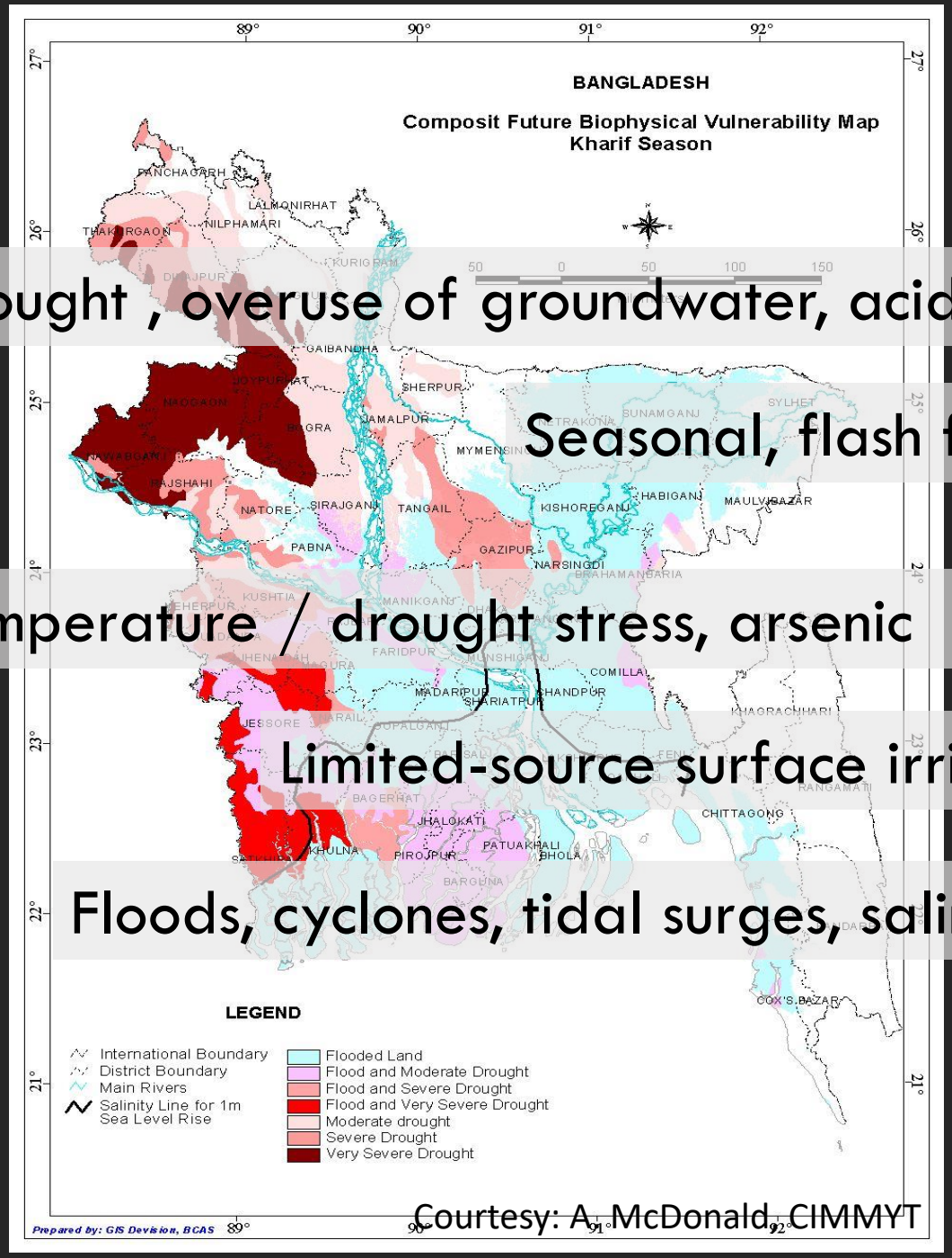
CGIAR – agricultural research for development

Climate Change Vulnerability Index 2017





R. Ortiz et al. http://www.ukm.my/ipi/wp-content/uploads/2013/07/35.2008Climate-changeCan-wheat-beat-the-heat.pdf



Drought , overuse of groundwater, acid soils

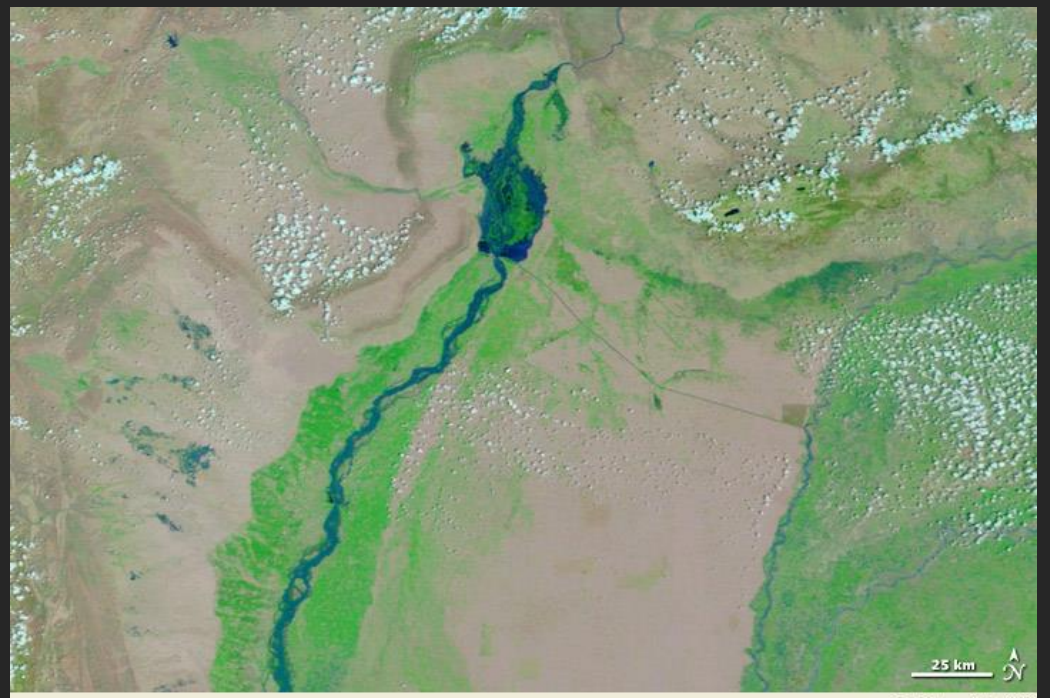
Seasonal, flash flooding

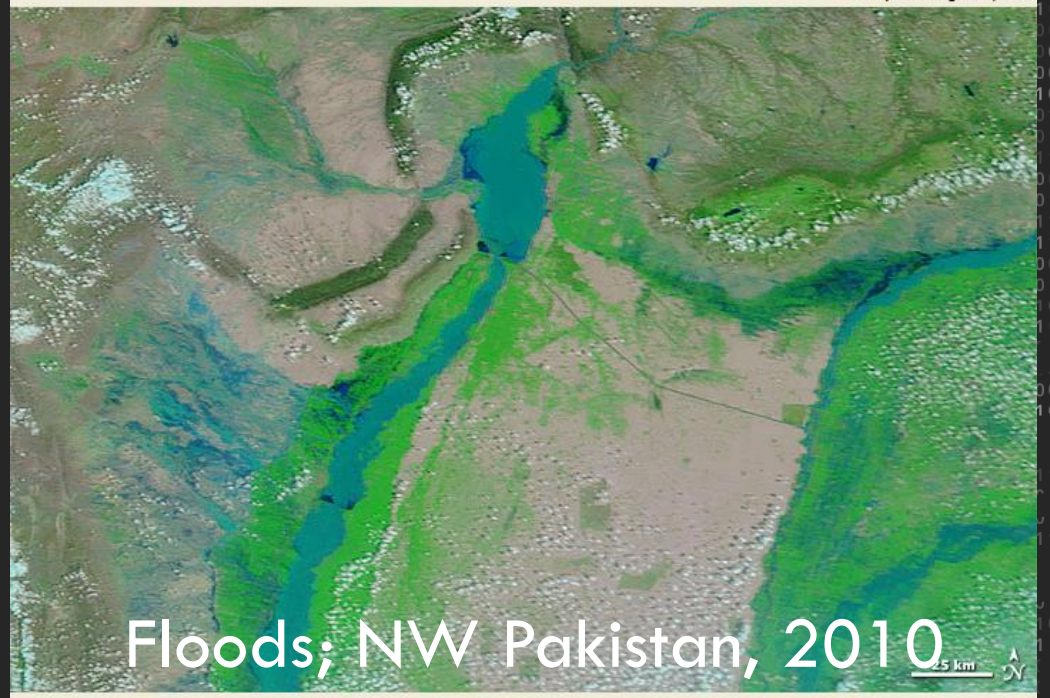
Temperature / drought stress, arsenic

Limited-source surface irrigation

Floods, cyclones, tidal surges, salinity

Courtesy: A. McDonald, CIMMYT





Floods; NW Pakistan, 2010

Ag R4D: Complex networks, systems, infrastructure…

Multidisciplinary (agronomy, breeding, socioeconomics, bioinformatics, data science)

Multi-scale (genetic/genomic to landscape)

Multi-stakeholder consultative, participatory processes (planning to implementing)

Highly heterogeneous, uncontrollable, challenging environments

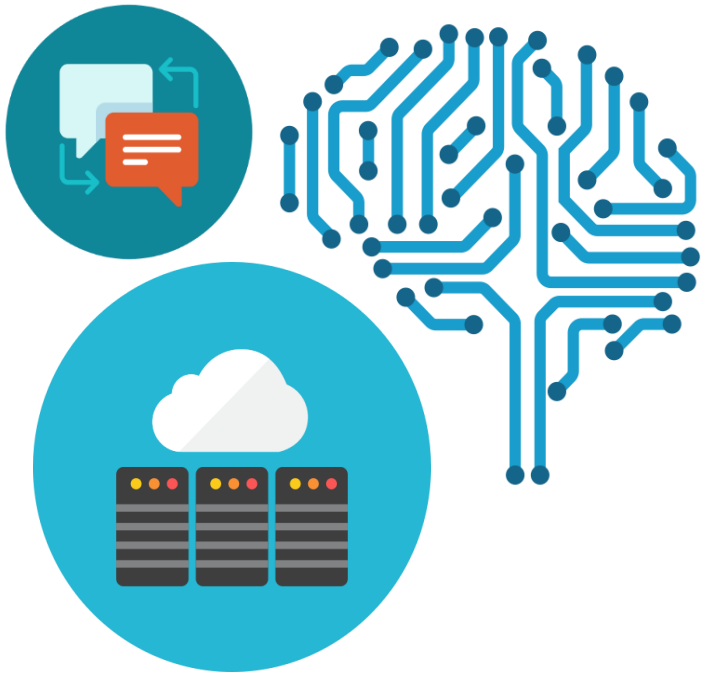
Hey Cigi, when should I plant my maize? How should I manage my crop?

Real-time **decision support** for farmers

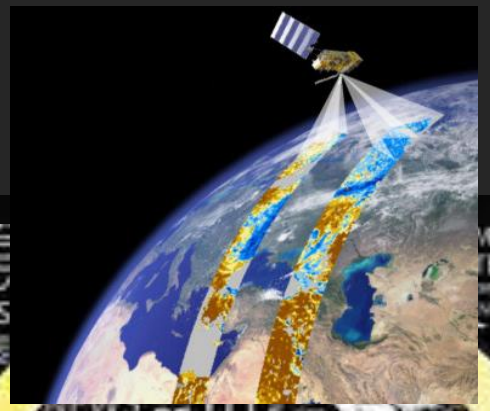
Easy **natural language** as an interface

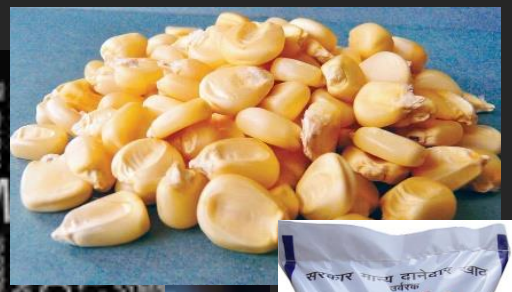
Smart **artificial intelligence** trained by CGIAR and partners

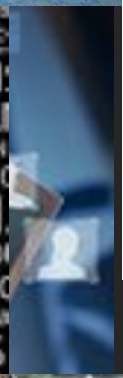
Leveraging multiple open, **harmonized and interoperable** databases



Opportunities







https://thelukewarmersway.wordpress.com/2016/02/07/climate-scientists-in-like-flint/

Making data: Findable; Accessible; Interoperable; Reusable

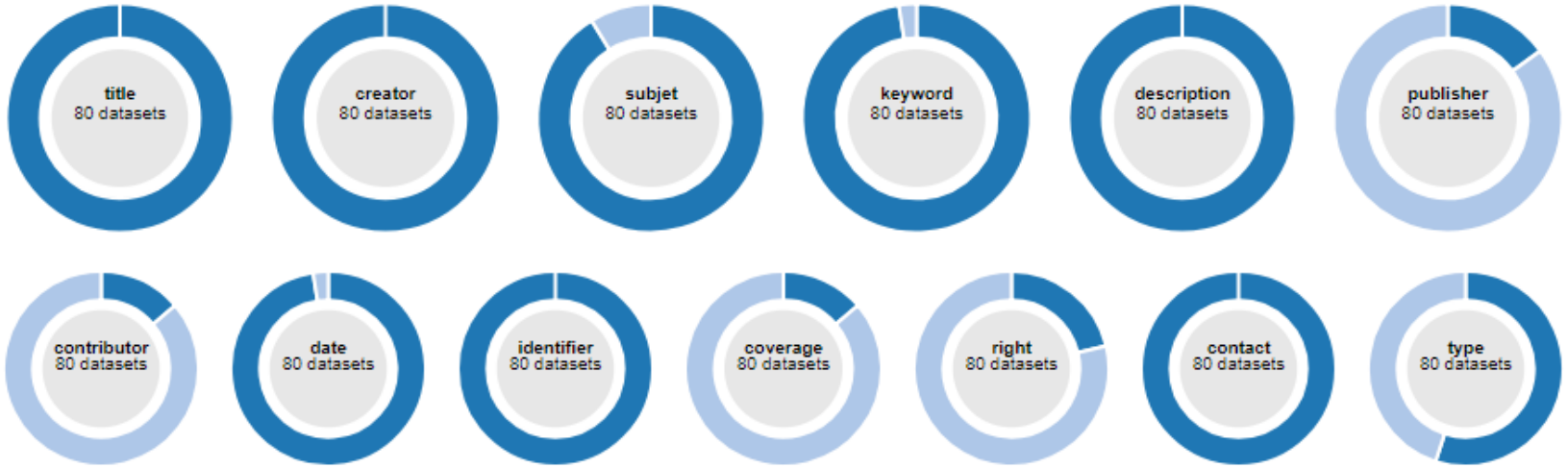
Data findability – CG Core Metadata Schema

DC Element	Qualifier	Required?	Definition
Title	Title of resource	Required	Official or unofficial title of the document, data set, image, etc.
Creator	Name of resource creator	Required	Creators of the item—typically a person. Could be an organization in case of corporate authors (e.g. Center reports)
Creator	ID of resource creator - if any	Required when applicable	ID of creator; use if ORCID, SCOPUS, or other type of creator ID scheme is in use
Creator	ID type of resource creator - if any	Required when applicable	Used to indicate the type of Creator ID – ex: SCOPUS, ORCID, etc.
Subject	General subject matter	Required	Subject matter of the research, technologies tested, etc.
Subject	AGROVOC subject term	Optional	AGROVOC subject matter or research area
Subject	Subject - other vocabularies (e.g. MeSH)	Required if applicable	Subject matter or research area from domain-specific vocabularies, if missing from AGROVOC
Description	Abstract of work	Required	Abstract or other description of the item
Publisher	Publisher of journal	Required when applicable	Entity responsible for publication, distribution, or imprint
Contributor	CGIAR Center name	Required	Research Centers with which creator(s) are affiliated
Contributor	non-CGIAR entity name	Required when applicable	Non-CGIAR partner entity with which creator/s are affiliated
Contributor	CRP	Required when applicable	CGIAR Research Program with which the research is affiliated
Contributor	Funding agency	Required	Funder, funding agency or sponsor
Contributor	Project	Required	Name of project with which the research is affiliated

Bioversity Dataverse

A total of 80 datasets has been uploaded on dataverse

percentage of datasets that comply with cg core metadata



Compliance to the cg core metadata per dataset

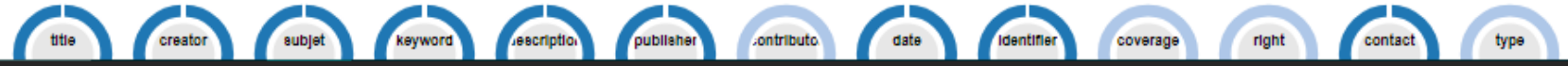
dataset http://hdl.handle.net/1902.1/21250

dataset http://dx.doi.org/10.7910/DVN/25450

dataset http://hdl.handle.net/1902.1/20804

dataset http://hdl.handle.net/1902.1/20803

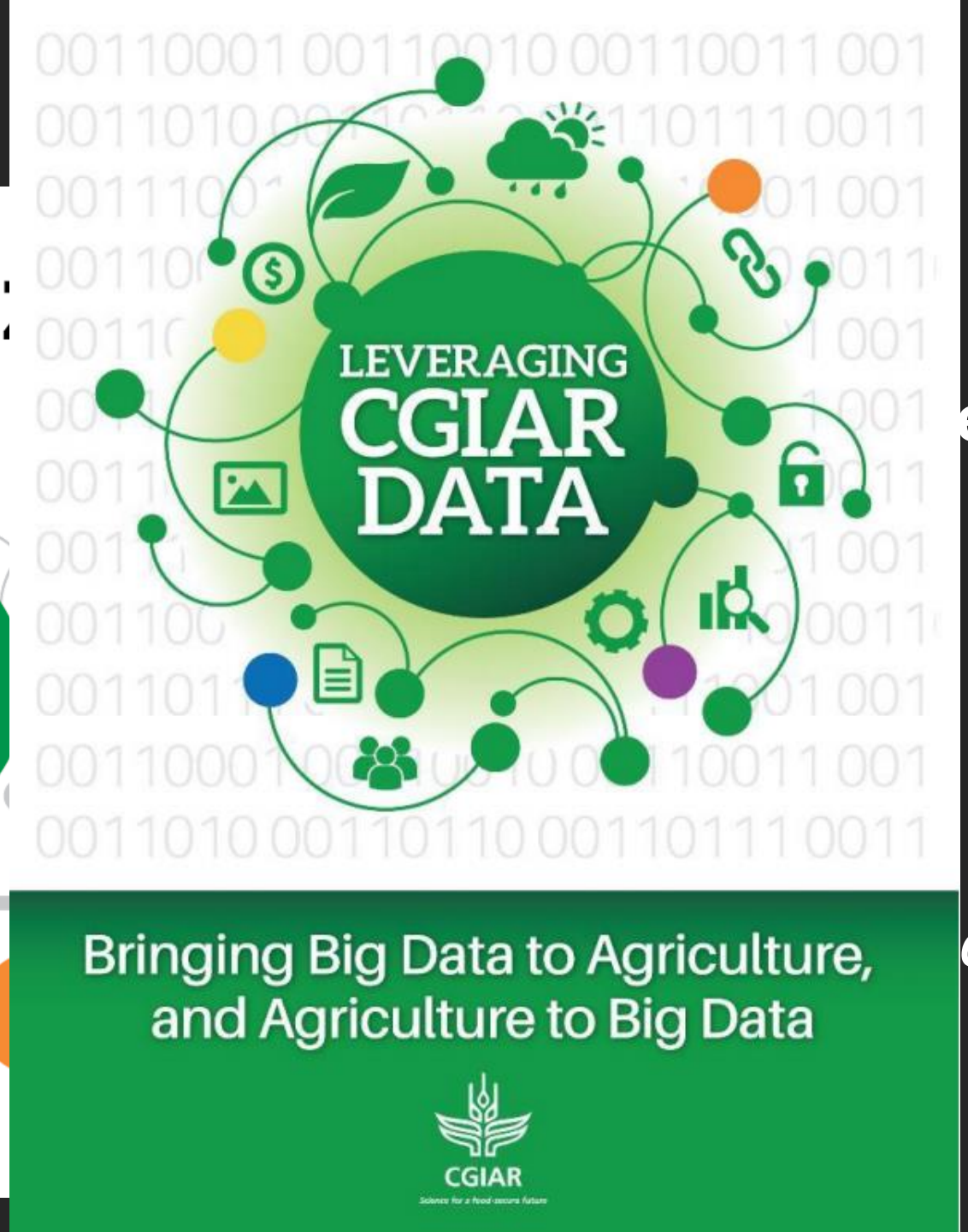
dataset http://dx.doi.org/10.7910/DVN/26086



(BIG) DATA HIGHWAY?

Opportunities

Organi



in data
ement

interoperability,

combining,
ol

Scale of Data

	Little Data	Shared Data	Aggregated Data	Aggregated or Combined, Depersonalized BIG DATA pool
Mary	my data	data shared	Aggregated co-created and connectable back to user	Aggregated or combined from multiple sources, de-personalized, but possibly connectable back to user
Bob	my data	data shared		
Sheri	my data	data shared		

Modified from: http://napsterization.org/stories/archives/cat_personal_data.html

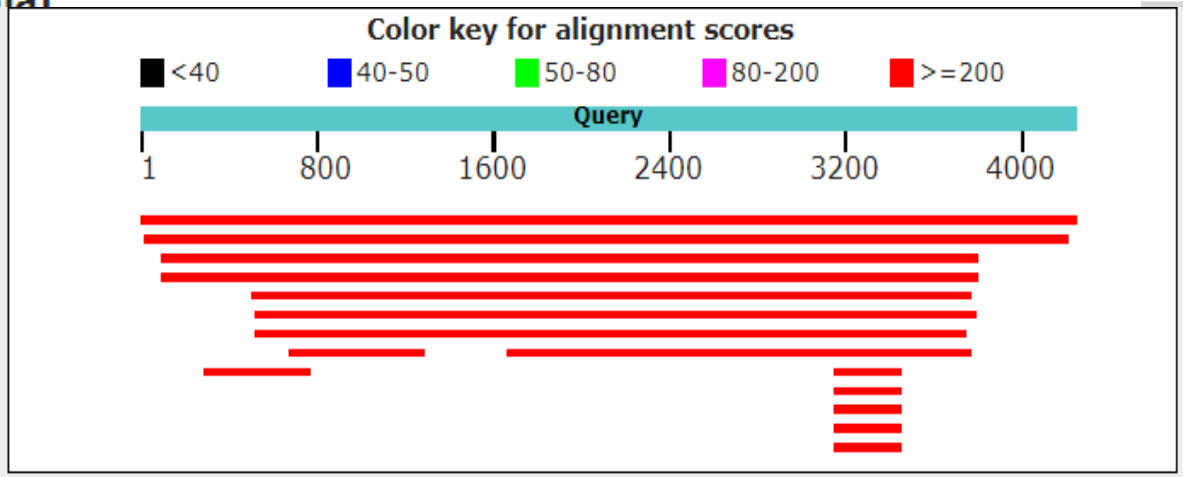
Approaches to data-driven innovation

Biomedical informatics → Only ~4% deposition into PubMed Central until mandated

Approaches to data-driven innovation

GenPept ▾ Send to: ▾ Change region shown

luciferase [Pyrocystis lunula]



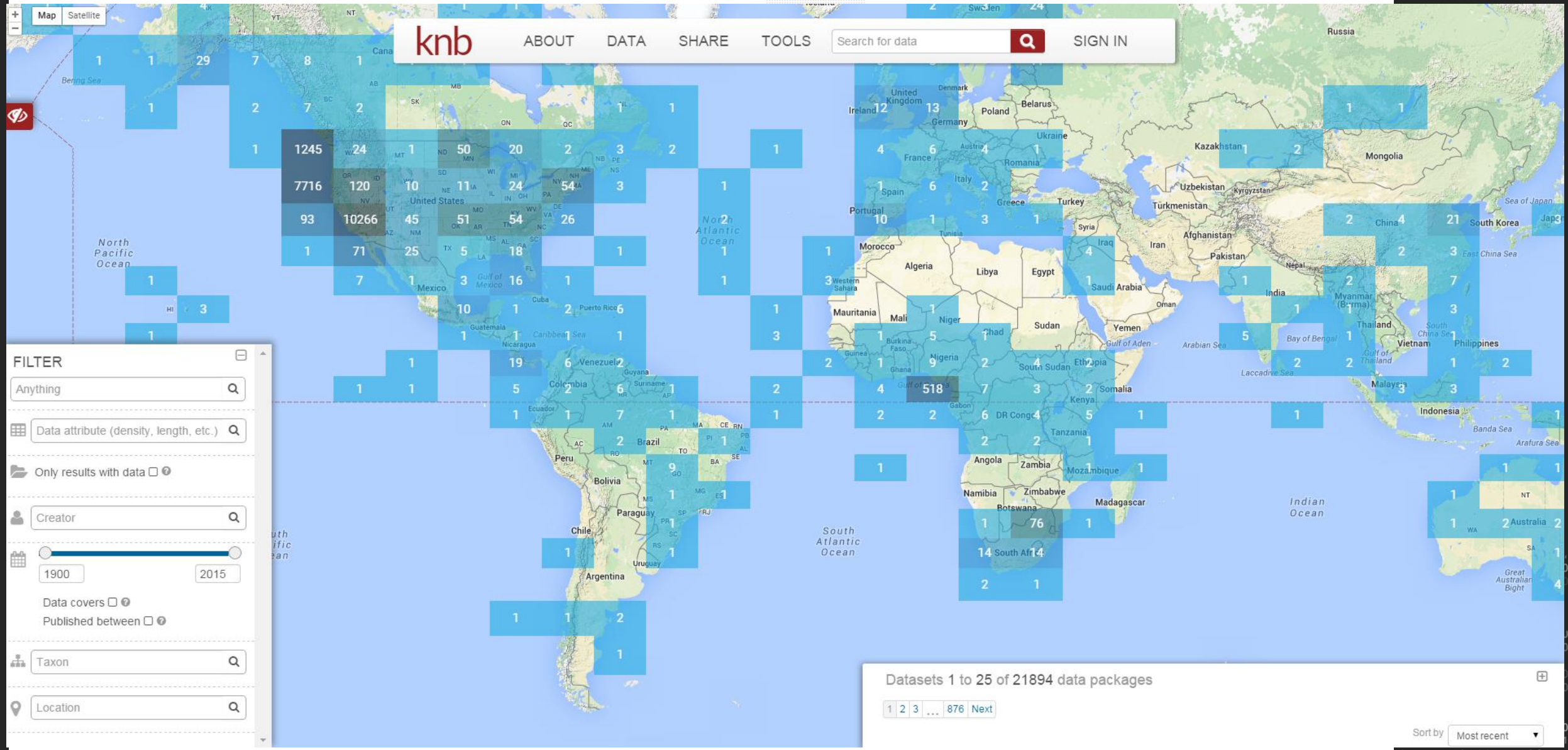
escriptions

Sequences producing significant alignments:

Select: All None Selected:0

Alignments Download GenBank Graphics Distance tree of results

Description	Max score	Total score	Query cover	E value	Ident	Accession
Pyrocystis lunula luciferase (lcfB) mRNA, complete cds	7808	9546	100%	0.0	100%	AF394060.1
Pyrocystis lunula luciferase (lcfA) mRNA, complete cds	6357	8027	98%	0.0	94%	AF394059.1
Pyrocystis fusiformis luciferase gene, complete cds	3648	4443	87%	0.0	85%	AY766384.1
Pyrocystis noctiluca luciferase gene, complete cds	3594	5216	87%	0.0	84%	AY766385.1
Gonyaulax polyedra luciferase mRNA, complete cds	2085	2085	77%	0.0	78%	AF085332.1
Alexandrium tamarense luciferase gene, complete cds	1783	1783	77%	0.0	77%	AY766383.1
Alexandrium affine luciferase gene, complete cds	1749	1749	76%	0.0	77%	AY766382.1


Map
Satellite
knb
ABOUT
DATA
SHARE
TOOLS
Search for data
SIGN IN
FILTER
Anything
Data attribute (density, length, etc.)
Only results with data
Creator
1900
2015
Data covers
Published between
Taxon
Location
Datasets 1 to 25 of 21894 data packages
1 2 3 ... 876 Next
Sort by
Most recent
LTER Network is not responsible for misinterpretation of data resulting from failure to consult metadata or data providers.

Getting to FAIR…

GARDIAN

HOME | SEARCH | ANALYTICS

DATA SEARCH

nutrition

All ■ Open □ Restricted □

And ■ Or □

SEARCH

Resources | Map | What If

Search results: **10106**

FILTERS

YEAR

All

CGIAR CENTER

All

LOOK IN

All fields

PUBLICATIONS (10023)

DATASETS (83)

GENETIC ACCESSIONS

COUNTRY All Countries GEOGRAPHIC SPREAD VIEW MAP

2014 **2014 Global Nutrition Report Dataset**
IFPRI

2017 **Food Trees Project Nutrition and Consumption Data**
ICRAF

2015 **2015 Global Nutrition Report Dataset**
IFPRI

2016 **Fruiting Africa Endline Consumption and Nutrition Survey**
ICRAF

2018 **His and Hers, time and income: How intra-household dynamics impact nutrition in agricultural households**
CIAT

2017 **Food Trees Project Data Collection Tools: Nutrition and Consumption**
ICRAF

2018 **Probabilistic Causal Models for Nutrition Outcomes of Agricultural Actions - Kenya model**
ICRAF

2013 **Integrating vegetables into maize-based systems for enhanced nutrition and income generation: Scoping study by AVRDC**
ILRI

2017 **Probabilistic Causal Models for Nutrition Outcomes of Agricultural Actions - Uganda model**
ICRAF

ACCESS RIGHTS

License

Not Defined

Terms of use

VIEW

LINKS

DOI

REFERS TO

DATASET

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

International Center for Tropical Agriculture (CIAT)

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

VIEW LARGE

FAIR COMPLIANCE

F = 4.75 / 5

A = 3.76 / 5

I = 3.60 / 5

R = 4.04 / 5

VIEW METRICS

DATASET FILES

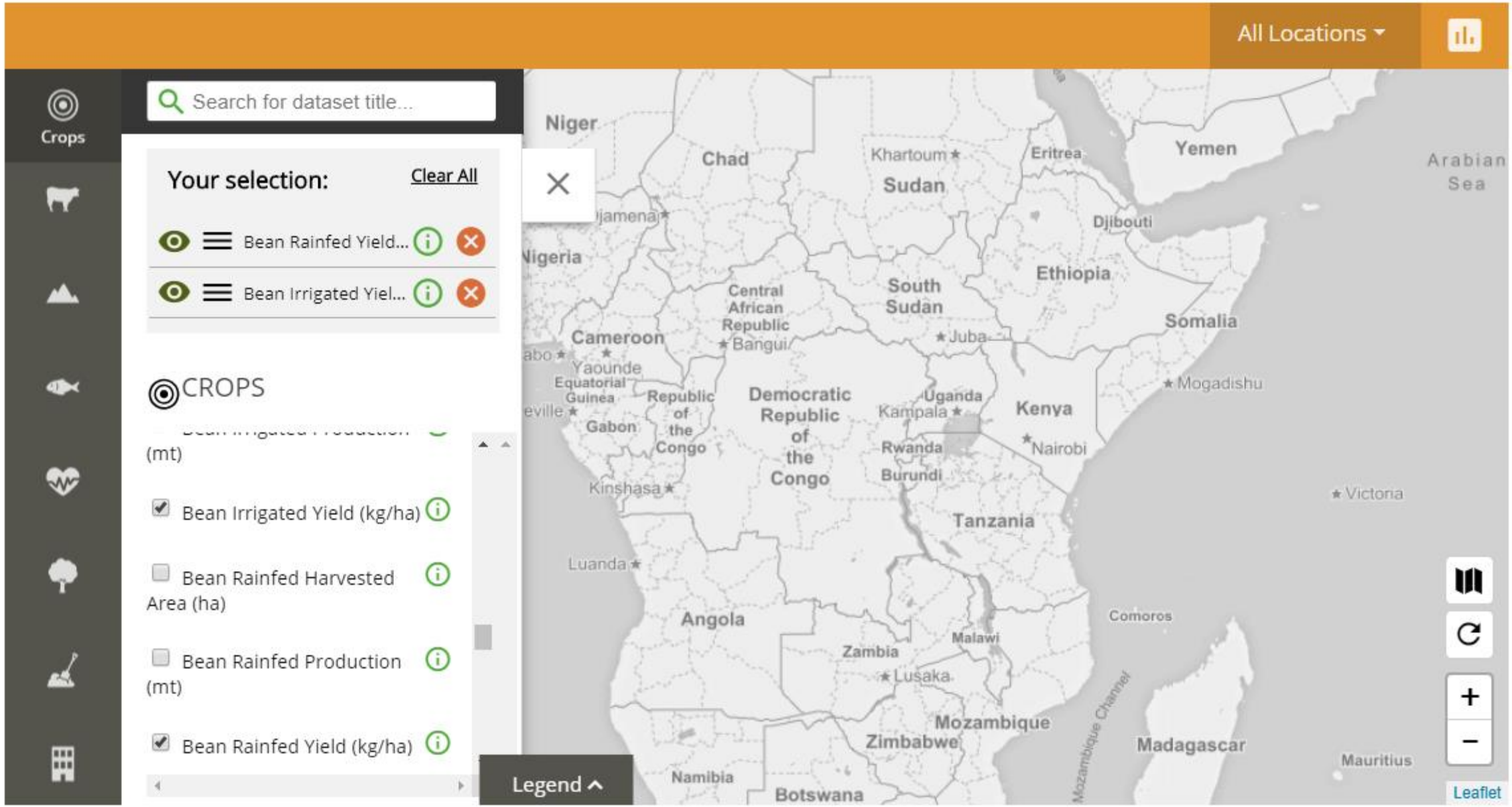
01. Household data collection Instrument.pdf
application/pdf

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 from South Africa

GEOSPATIAL DATA



GARDIAN

Powered by Agroknow

- Search
- Upload
- Map
- metaData API
- What If
- Statistics

Select a Question or make your own using SPARQL:

Select a Question

Select a Question

1. Give me all maize yields for Tanzania under no fertilizer and complete fertilizer

2. Give me yields of improved and not-improved maize in Long and Matufa

3. Give me average yields of improved and not improved maize reported by households along with average C and N content in the topsoil in Long and Matufa

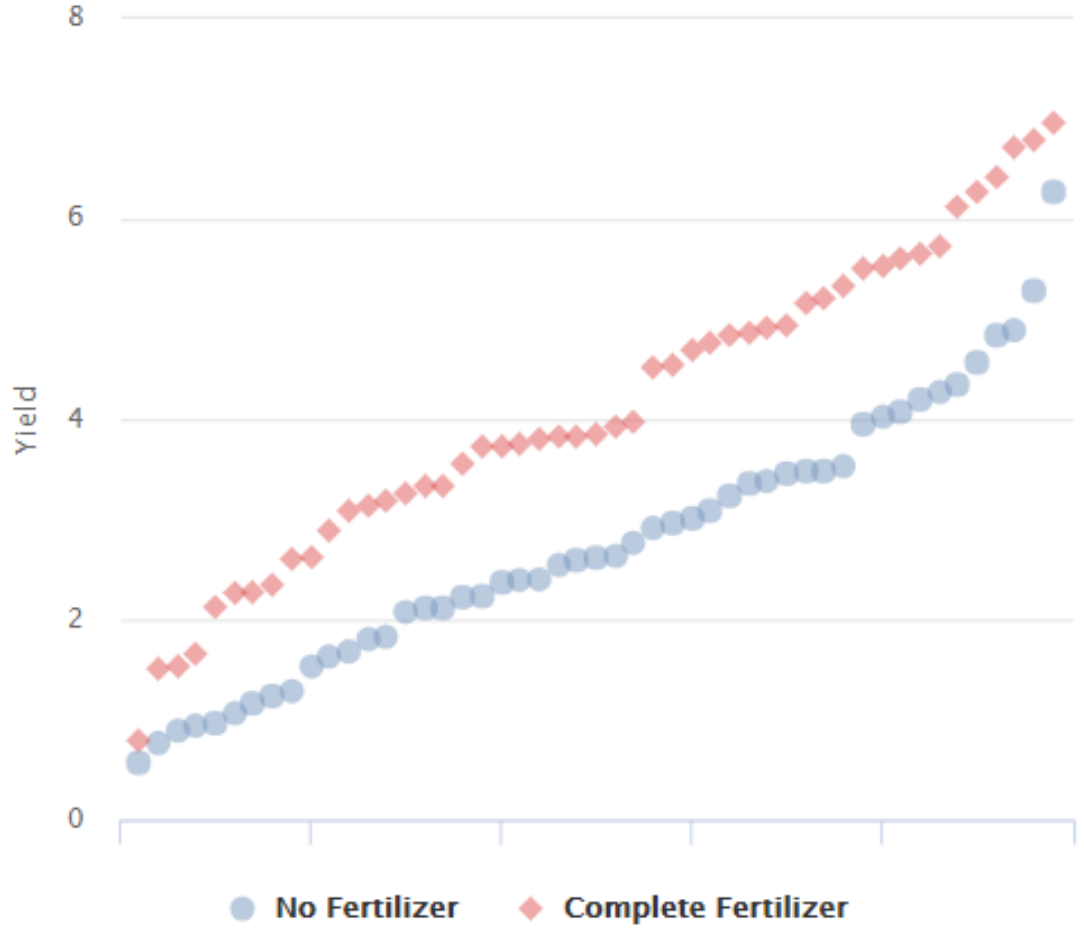
1. Give me all maize yields for Tanzania under no fertilizer and complete fertilizer

SPARQL Query

Execute

```
PREFIX cgiar: <http://data.cgiar.org/demo#> select DISTINCT ?t ?y ?tr where { ?t cgiar:yield ?y .  ?t cgiar:inZone ?z . ?t cgiar:treatment ?tr . {?t cgiar:treatment <http://data.cgiar.org/demo/resource/treatment/NPK>} UNION {?t cgiar:treatment <http://data.cgiar.org/demo/resource/treatment/Control>} . ?z cgiar:belongsTo <http://data.cgiar.org/demo/resource/country/Tanzania> .}
```

Result



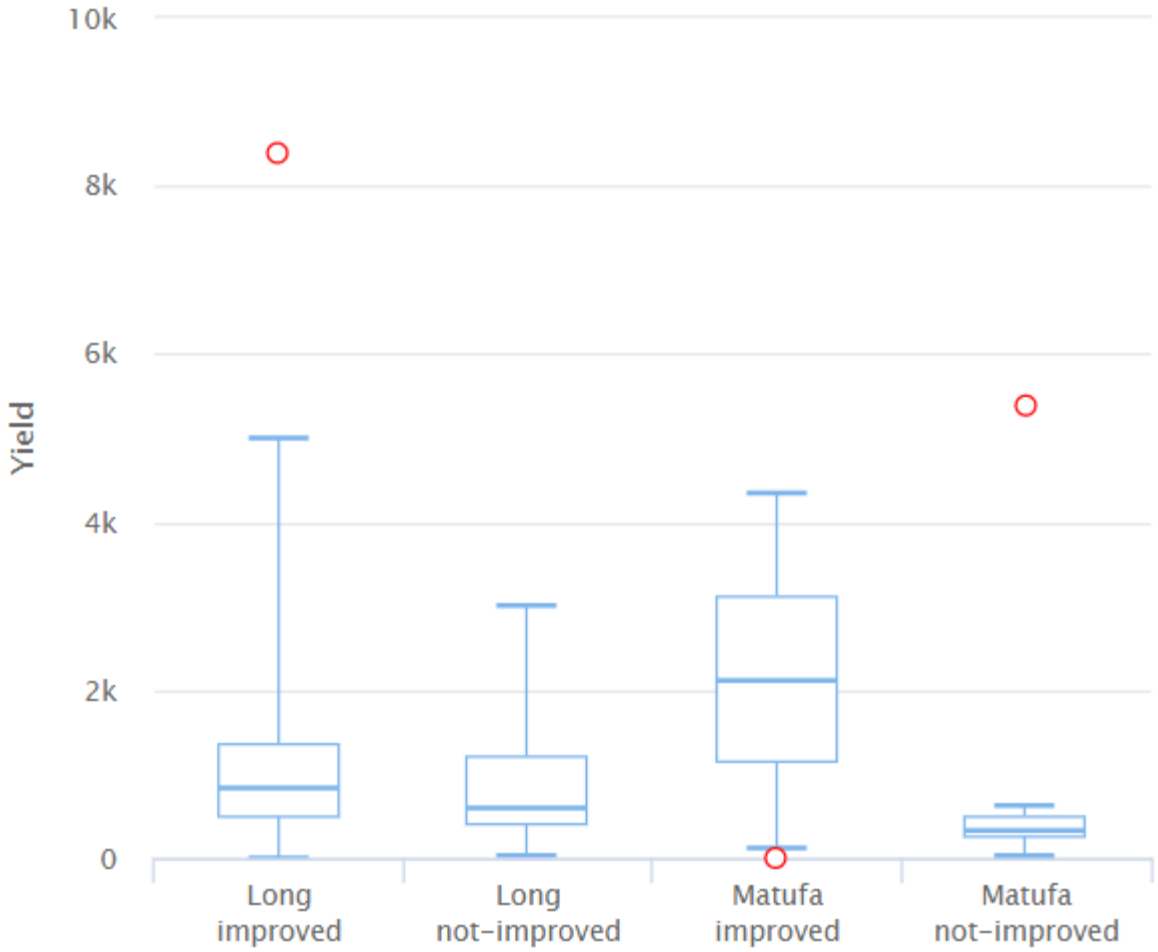
2. Give me yields of improved and not-improved maize in Long and Matufa

SPARQL Query

Execute

```
PREFIX cgiar: <http://data.cgiar.org/demo#> select ?r ?kind ?yield
where { ?s cgiar:onVillage ?r ; cgiar:yield ?yield ;
cgiar:improvedCrop ?kind . } ORDER BY ?r ?kind
```

Result



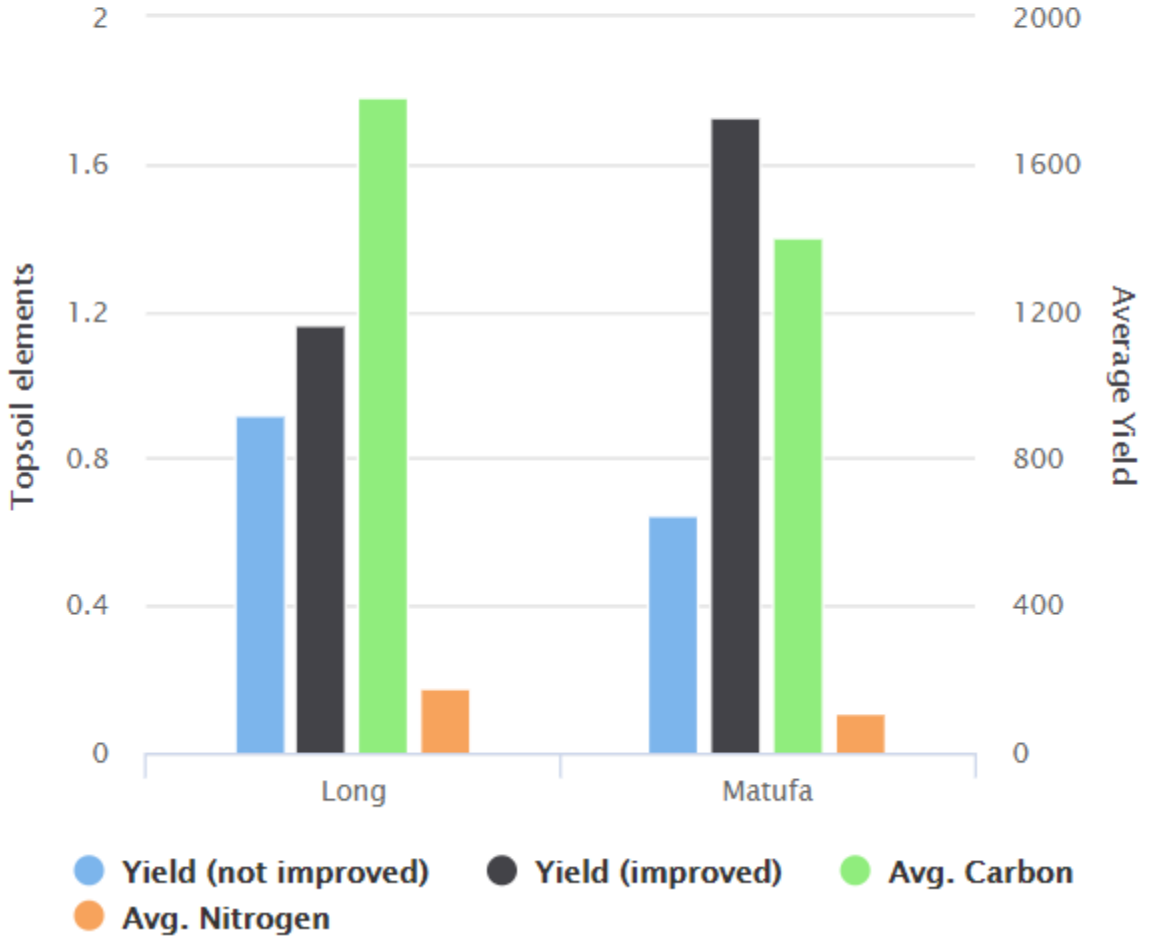
3. Give me average yields of improved and not improved maize reported by households along with average C and N content in the...

SPARQL Query

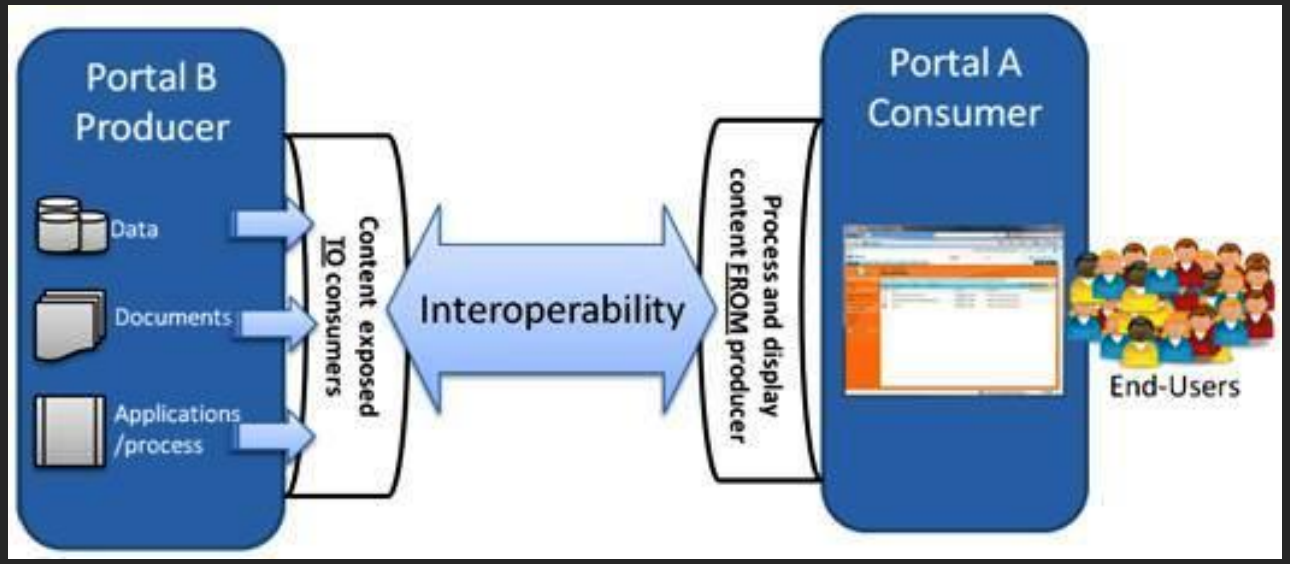
Execute

```
PREFIX cgiar: <http://data.cgiar.org/demo#> select ?r (AVG(?improved) as ?improved_yield) (AVG(?not_improved) as ?not_improved_yield) (AVG(?c) as ?carbon) (AVG(?n) as ?nitrogen) where { ?t cgiar:hasDepthCode "Topsoil" ;  cgiar:onSite ?r ; cgiar:carbon ?c ;  cgiar:nitrogen ?n .  ?s cgiar:onVillage ?r ; cgiar:yield ?improved ;  cgiar:improvedCrop "1"^^xsd:integer . ?m cgiar:onVillage ?r ;  cgiar:yield ?not_improved ; cgiar:improvedCrop "2"^^xsd:integer .} GROUP BY ?r ORDER BY ?r
```

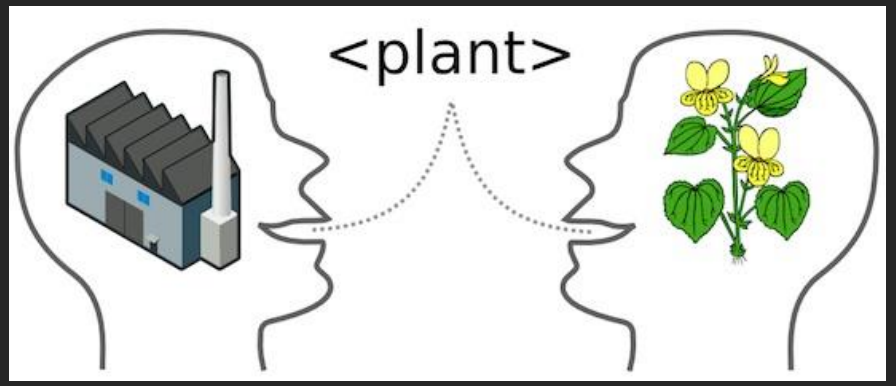
Result



Interoperability



Syntactic interoperability: machines communicate and exchange data



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

Traits, methods and scales

DOWNLOAD | SHOW OBSOLETE TERMS

Rice trai
Abiot
al
co
dr
dr
he
irc
ph
sa
su
zir
Agron
Bioch

Vereinigtes Königreich

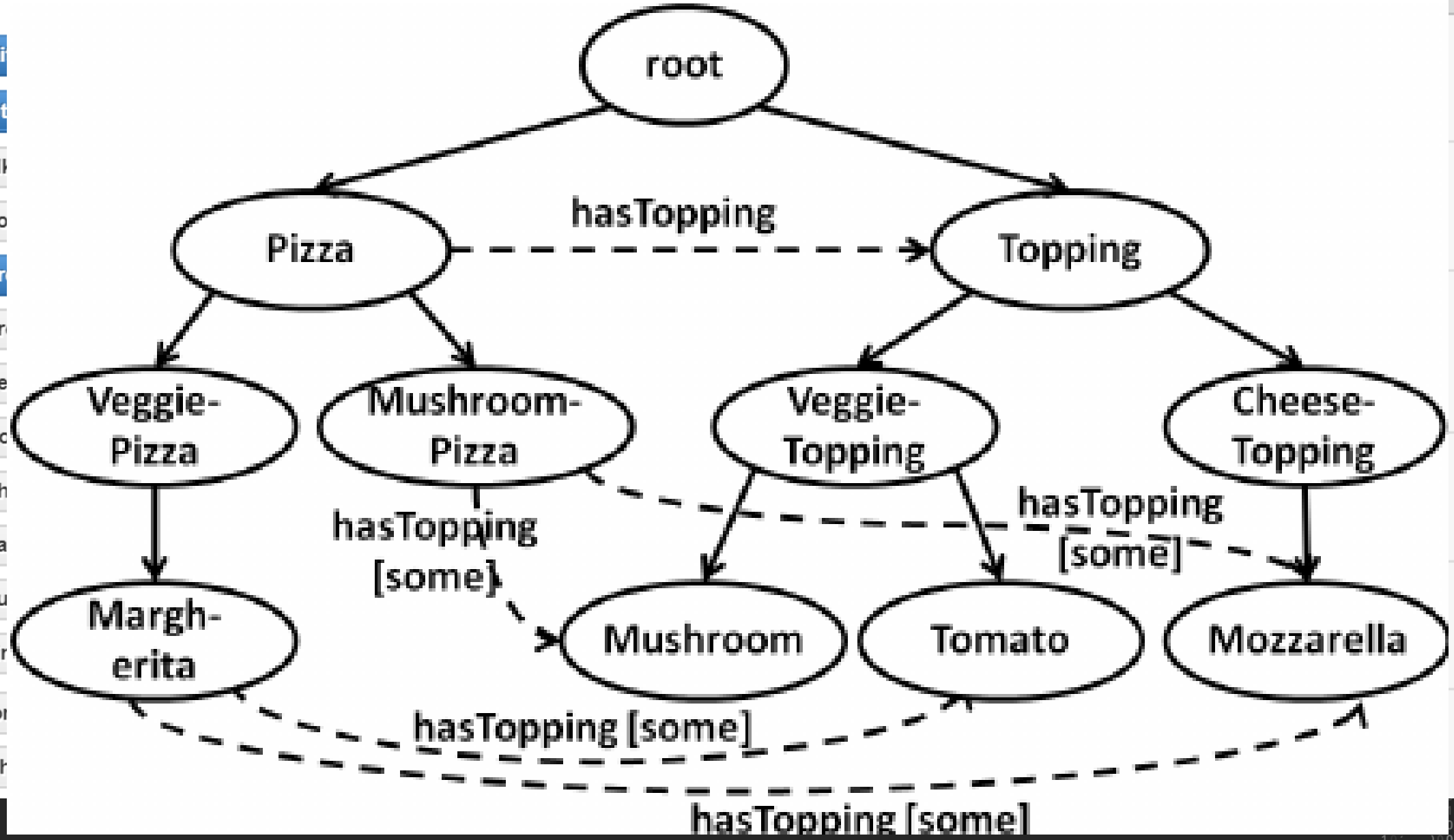
Earth (1)

parent

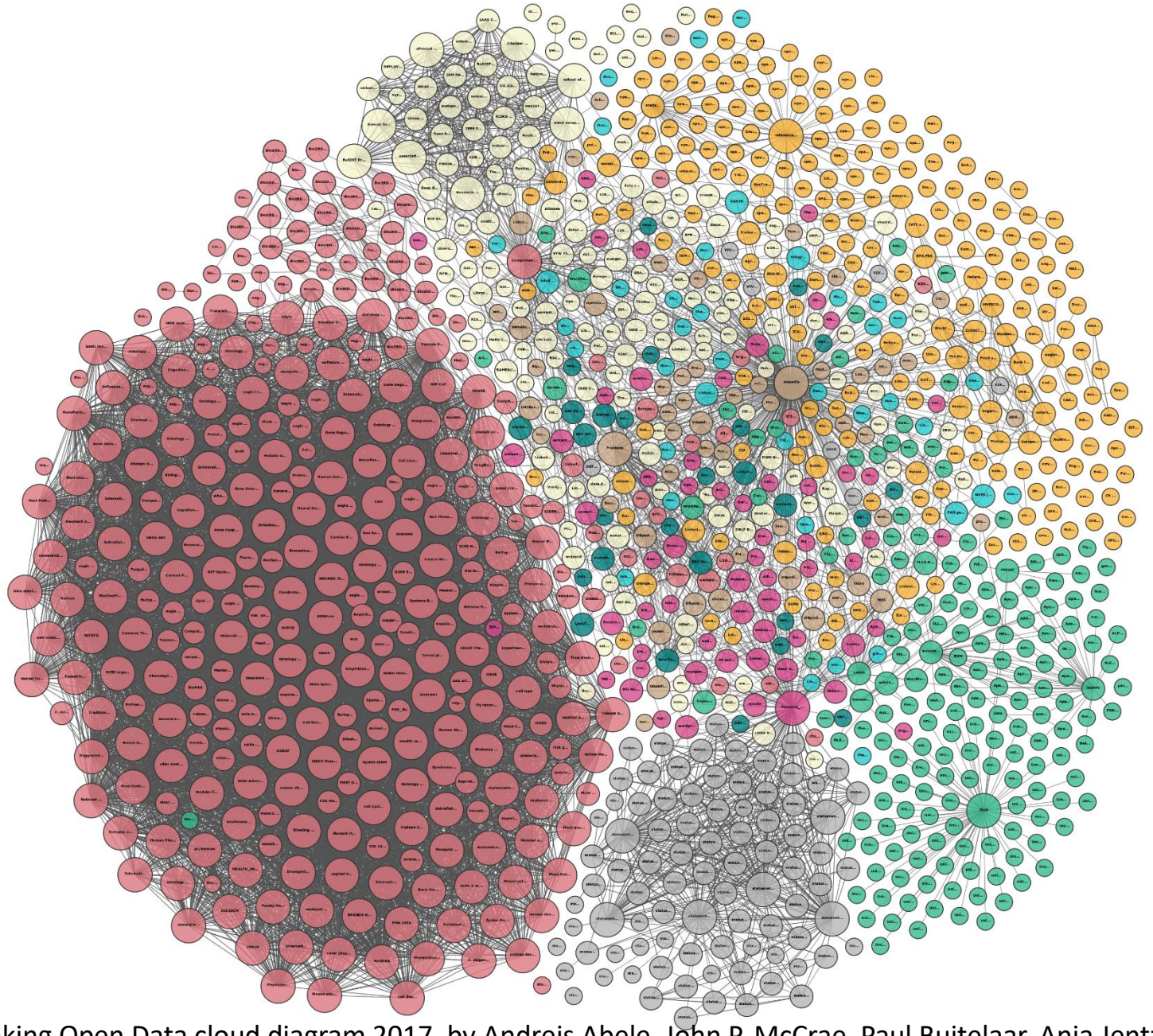
Crop Ontology
for agricultural data

ables. Click on a term to display

0 Comments



Interoperability – Linked Open Data



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

HIDAP AgroFIMS

Hello, Guest
Not connected
Log in

- Fieldbook
- Single Trial Analysis
- Documentation
- Help
- About

Platform for Big Data in Agriculture

HIDAP AgroFIMS

Agronomy Field Management System

HIDAP AgroFIMS v0.0.17

The Agronomy Field Information Management System (AgroFIMS) has been developed on CGIAR's HiDAP (Highly-interactive Data Analysis Platform created by CGIAR's International Potato Center, CIP). AgroFIMS draws fully on ontologies, particularly the Agronomy Ontology and the Crop Ontology. It consists of modules that represent the typical cycle of operations in agronomic trial management, and enables the creation of data collection sheets using the same ontology-based set of variables, terminology, units and protocols. AgroFIMS therefore:

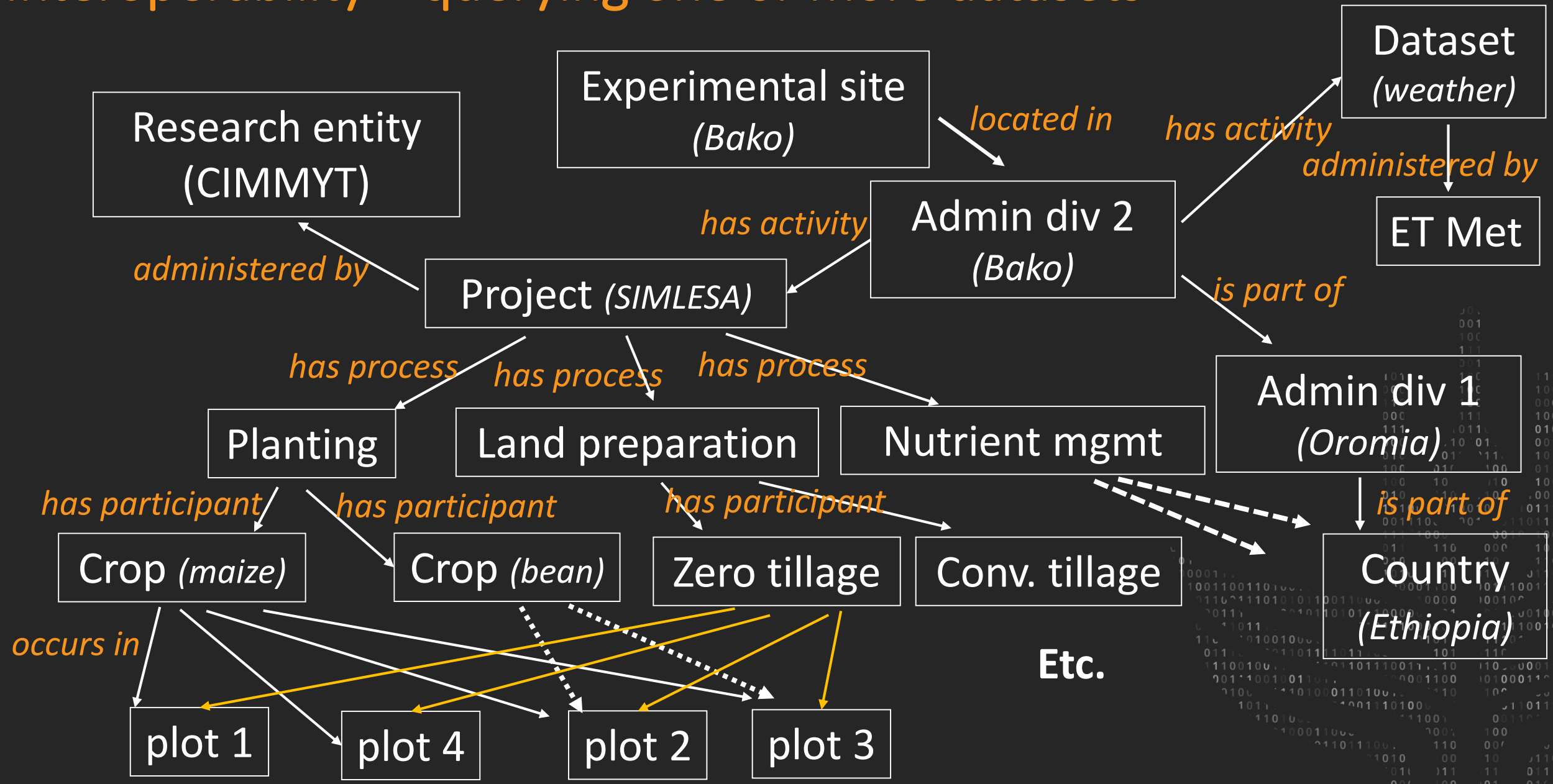
- Standardizes data collection and description for easy aggregation and inter-linking across disparate datasets;
- Allows easy integration with HiDAP breeding data, or any other ontology-based datasets;
- Functions as a data staging repository, allowing data uploads with view/edit permissions;
- Enables data quality checks, statistical analysis of the data collected, and the generation of sophisticated statistics reports;
- Aligns a priori with CGIAR's CG Core metadata schema;
- Enables easy upload to the institutional repositories, and much more.

Funding for AgroFIMS was provided by the Bill and Melinda Gates Foundation's Open Access, Open Data Initiative, and the CGIAR Big Data Platform.

	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 afte	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
		Phosphorus content of DAP basal fertilizer	Numeric	kg/ha	36.5838	46	0	69	24.2365779	226
	DAP (P&N)	Nitrogen content of Diammonium phosphate fertilizer	Numeric	kg/ha						
		Phosphorus content ofapplied NPK basal fertilizer	Numeric	kg/ha						
	NPK	Nitrogen content of applied NPK fertilizer	Numeric	kg/ha						
	Urea	Nitrogen content of applied Urea fertilizer	Numeric	kg/ha						

Mozamb	Cabango	159	3	Farmers check	SOLE	BASIN	CA	2013-14		984.9	382.4					
Mozamb	Cabango	157	4	Basins Maize sole	SOLE	FLAT	CP	2013-14		1024.3	346					
Mozamb	Cabango	157	4	jab planter +R/up+Maize sole	SOLE	BASIN	CA	2013-14		1024.3	346					
Mozamb	Cabango	157	4	jab planter Maize-common bean rota	ROTCP	FLAT	CA	2013-14		1024.3	346					
Mozamb	Cabango	157	4	jab planter Maize-common bean Inte	INTCP	BASIN	CA	2013-14		1024.3	346					
Mozamb	Cabango	157	4	Farmers check	SOLE	BASIN	CA	2013-14		1024.3	346					
Mozamb	Cabango	439	5	Basins Maize sole	SOLE	FLAT	CP	2013-14		618.3	85					
Mozamb	Cabango	439	5	jab planter +R/up+Maize sole	SOLE	BASIN	CA	2013-14		618.3	85					
Mozamb	Cabango	439	5	jab planter Maize-common bean rota	ROTCP	FLAT	CA	2013-14		618.3	85					
Mozamb	Cabango	439	5	jab planter Maize-common bean Inte	INTCP	BASIN	CA	2013-14		618.3	85					
Mozamb	Cabango	439	5	Farmers check	SOLE	BASIN	CA	2013-14		618.3	85					
Mozamb	Cabango	257	6	Basins Maize sole	SOLE	FLAT	CP	2013-14		930.7	67					
Mozamb	Cabango	257	6	jab planter +R/up+Maize sole	SOLE	BASIN	CA	2013-14		930.7	67					
Mozamb	Cabango	257	6	jab planter Maize-common bean rota	ROTCP	FLAT	CA	2013-14		930.7	67					
Mozamb	Cabango	257	6	jab planter Maize-common bean Inte	INTCP	BASIN	CA	2013-14		930.7	67					
Mozamb	Cabango	257	6	Testemunha do campones	SOLE	BASIN	CA	2013-14		930.7	67					
Mozamb	Sussundenga	165	1	Covacho Milho puro	SOLE	FLAT	CP	2013-14		1196	471.4					
Mozamb	Sussundenga	165	1	Matraca +R/up+Milho puro	SOLE	BASIN	CA	2013-14		1196	471.4					
Mozamb	Sussundenga	165	1	Matraca Feijao vulgar-Milho rotacac	ROTCP	FLAT	CA	2013-14		1196	100.8					
Mozamb	Sussundenga	165	1	Matraca Milho-Feijao vulgar consoc	INTCP	BASIN	CA	2013-14		1196	471.4					
Mozamb	Sussundenga	165	1	Testemunha do campones	SOLE	BASIN	CA	2013-14		1196	471.4					
Mozamb	Sussundenga	342	2	Covacho Milho puro	SOLE	FLAT	CP	2013-14		1493.6	471					
Mozamb	Sussundenga	342	2	Matraca +R/up+Milho puro	SOLE	BASIN	CA	2013-14		1493.6	471					

Interoperability – querying one or more datasets



Let’s explore SIMLESA data!

What does this mean for you?

What lies ‘beyond metadata’ or business as usual to leverage data science and analytics capabilities and needs fully?

e.g. sample DoIs; interfacing physical with digital? Hi-throuput support?

Can we build tools to ease metadata entry to repositories/dbs? To semantically enrich datasets?

How might libraries, librarians/info-data specialists evolve roles?

Thank you!

Platform for
Big Data
in Agriculture

bigdata.cgiar.org