Applying FAIR principles to CGIAR data

Making data reusable for models and other quantitative analyses

Cheryl Porter

presented to the

Big Data in Agriculture Course

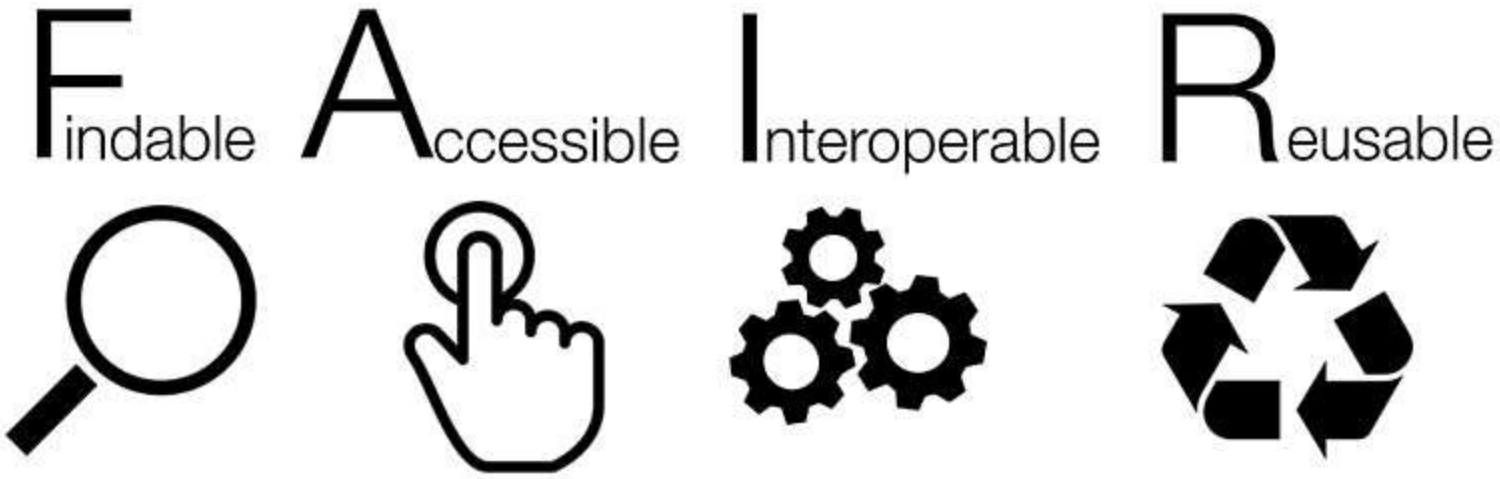
Rabat, Morocco

2018-12-10

UF | UNIVERSITY of FLORIDA

AgMIP

Open Data – FAIR principles



https://www.force11.org/group/fairgroup/fairprinciples

Why should I share my data?

- Increased efficiency – no need to duplicate experiments
- Research citations for your data contributions
- Promote your research
- Allows data to be combined with data of other types for novel new research
- Mandates from funding sources, journals, governments

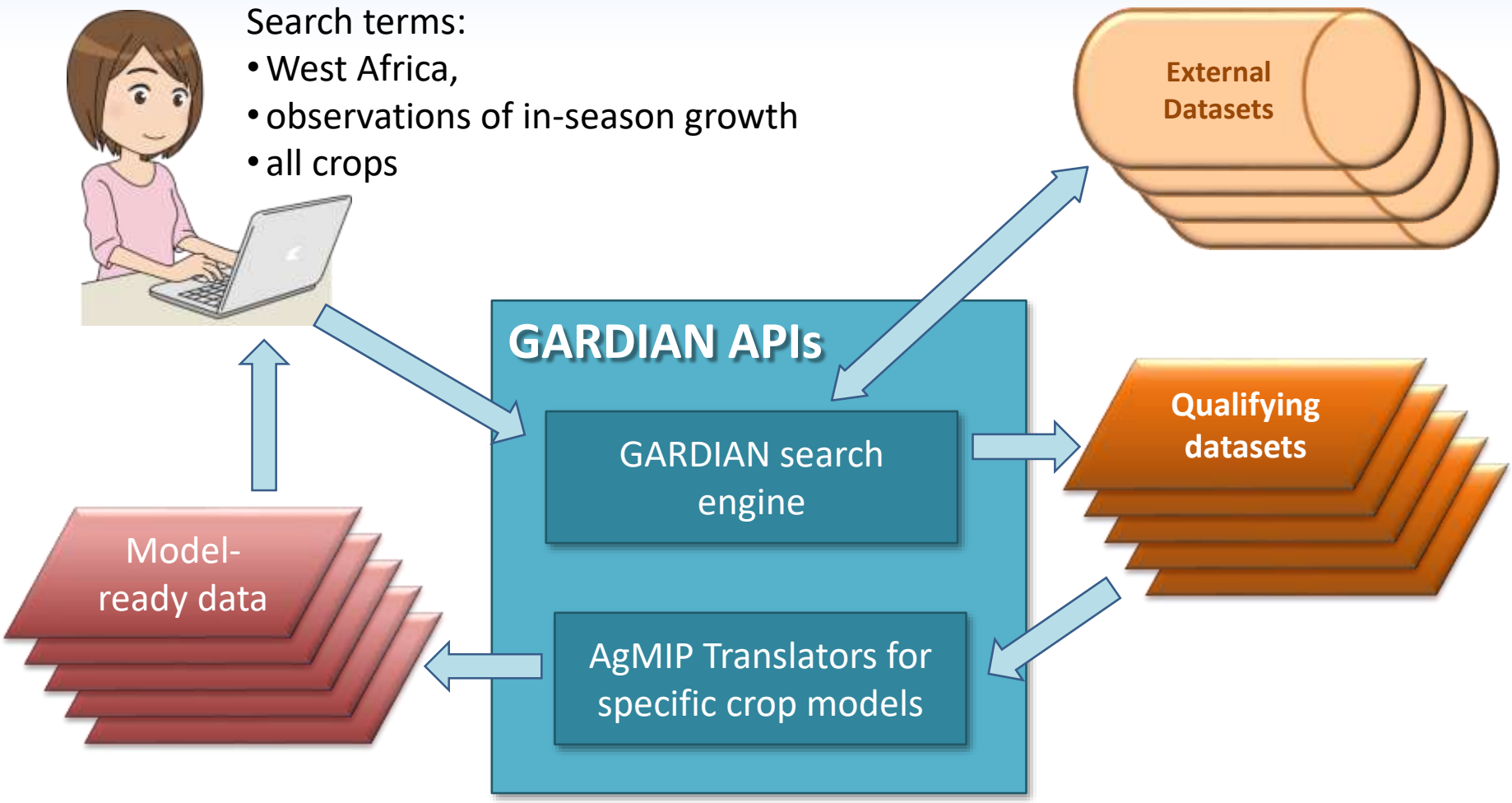
Quantitative re-use of agronomic data:

- meta-analyses
- evidence-based decision support
- modeling
- cross-disciplinary research

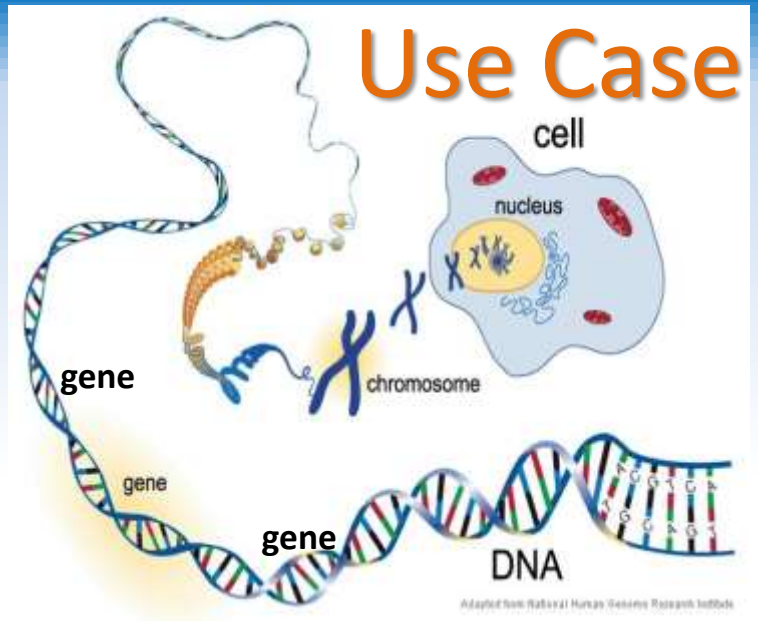
from: www.wholewhale.com

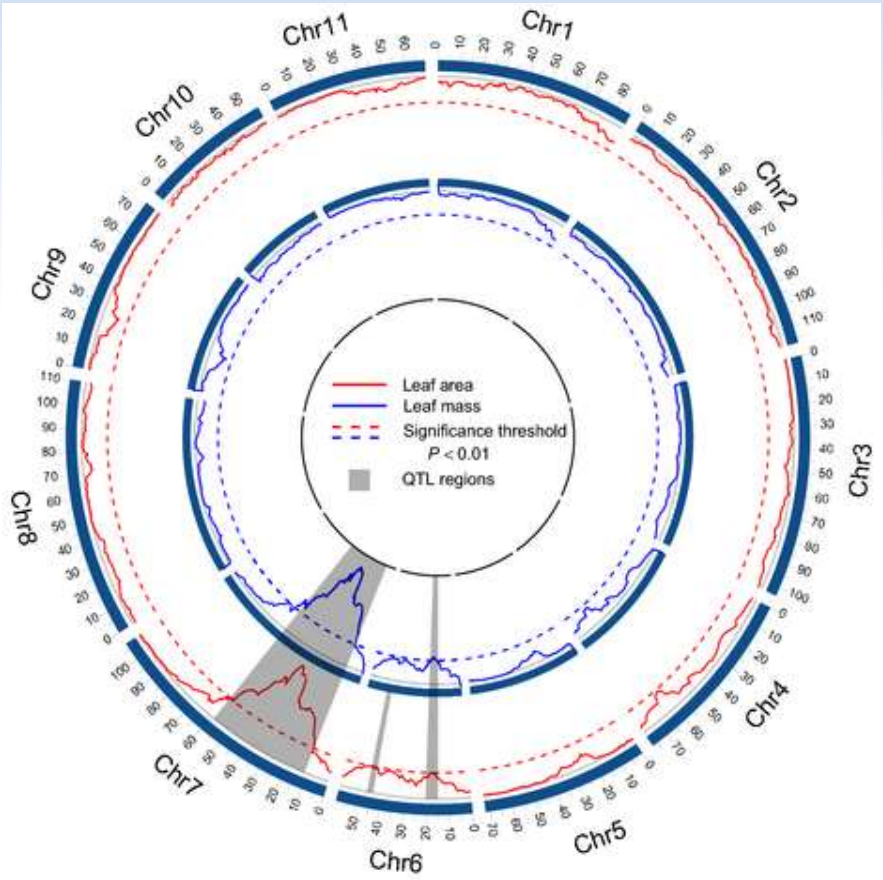
Use Case 1 – Crop modeler

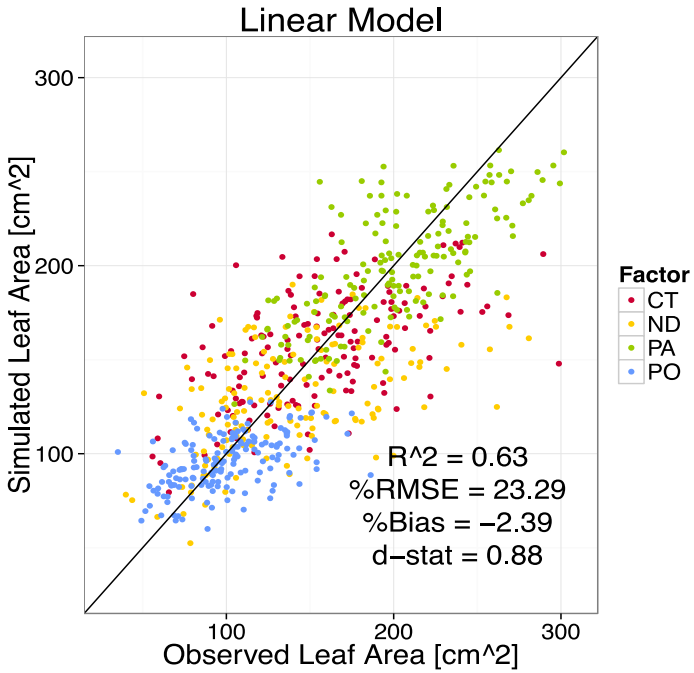
Ava, a crop modeler, is developing, evaluating and recommending sustainable intensification practices in West Africa.



Use Case 2 - Gene-based modeling

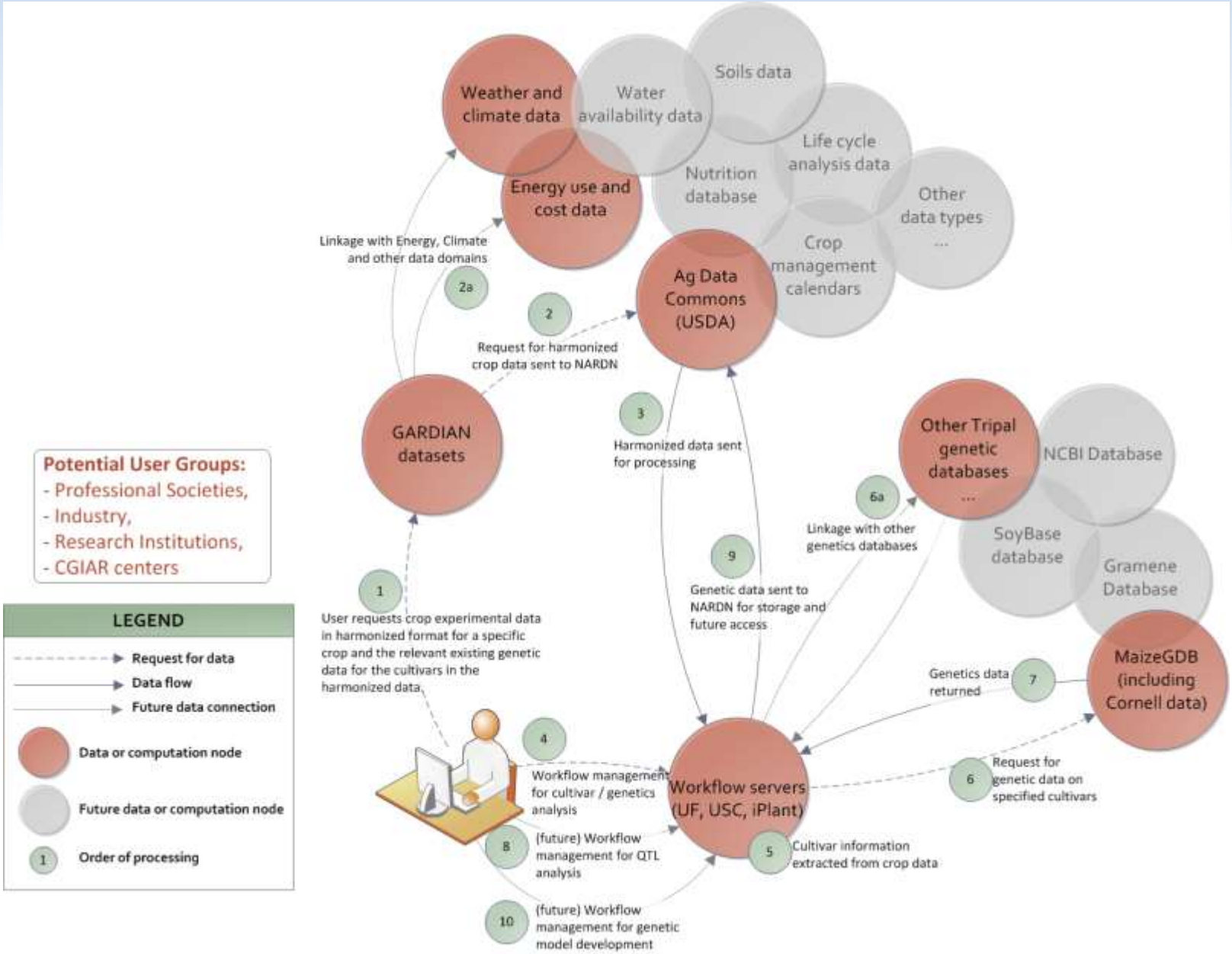






$$MSLAmax = \overline{MSLAmax} + a_i \cdot (mean(E_i) - \bar{E}_i) + b_j \cdot QTL_j + c_{ij} \cdot \left(QTL_j \cdot (mean(E_i) - \bar{E}_i)\right)$$

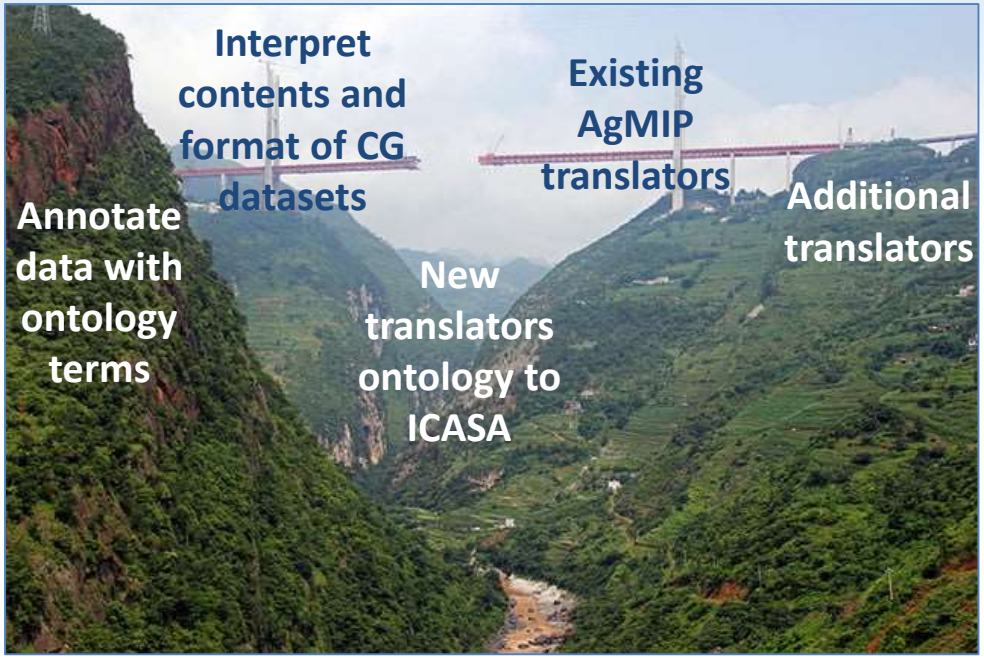
Network of Networks



Using AgMIP data interoperability tools to "Bridge the Gap"

Data Supply

- Multiple databases,
- research experiments,
- farm surveys,
- government data,
- NGO data,
- etc...



Data Demand

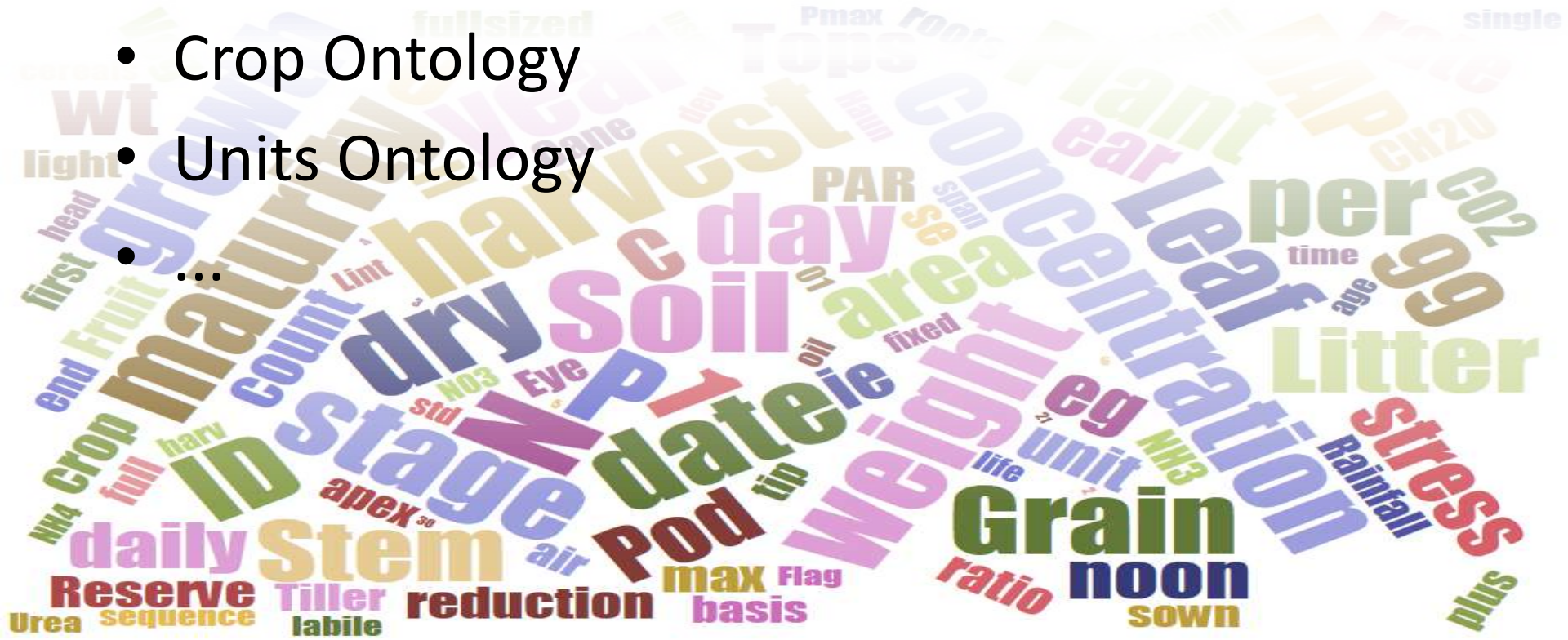
- Decision support,
- meta-analyses,
- cross-domain analyses,
- modeling,
- data analytics,
- etc...

Challenges

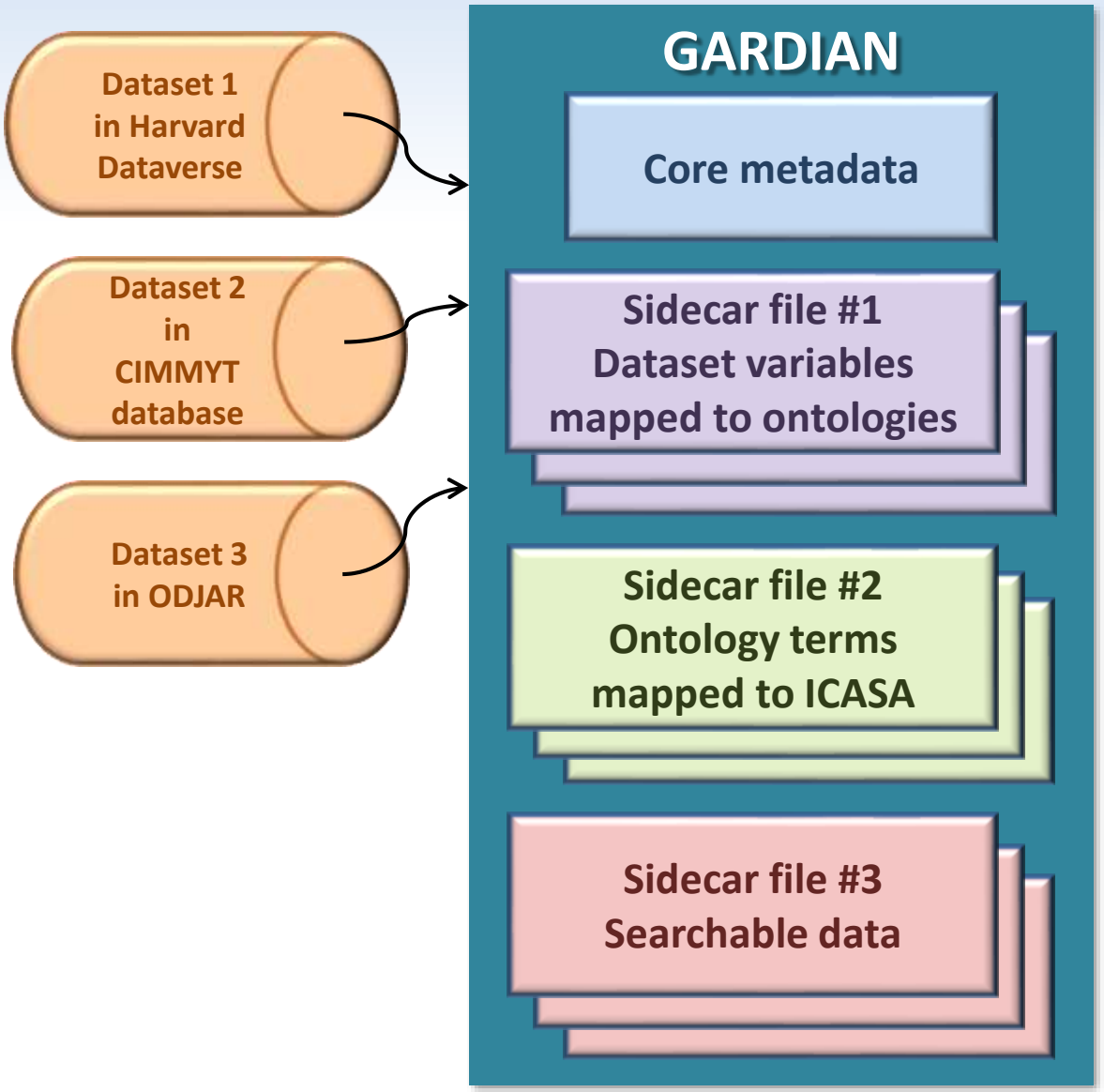
- Formats,
- Database schemas,
- locations,
- domains,
- vocabularies,
- ...

Standardized Vocabularies

- ICASA – International Consortium for Agricultural Systems Applications
- AgrO – Agronomy Ontology
- Crop Ontology
- Units Ontology
- …



Annotation to allow search, access, and translation of data



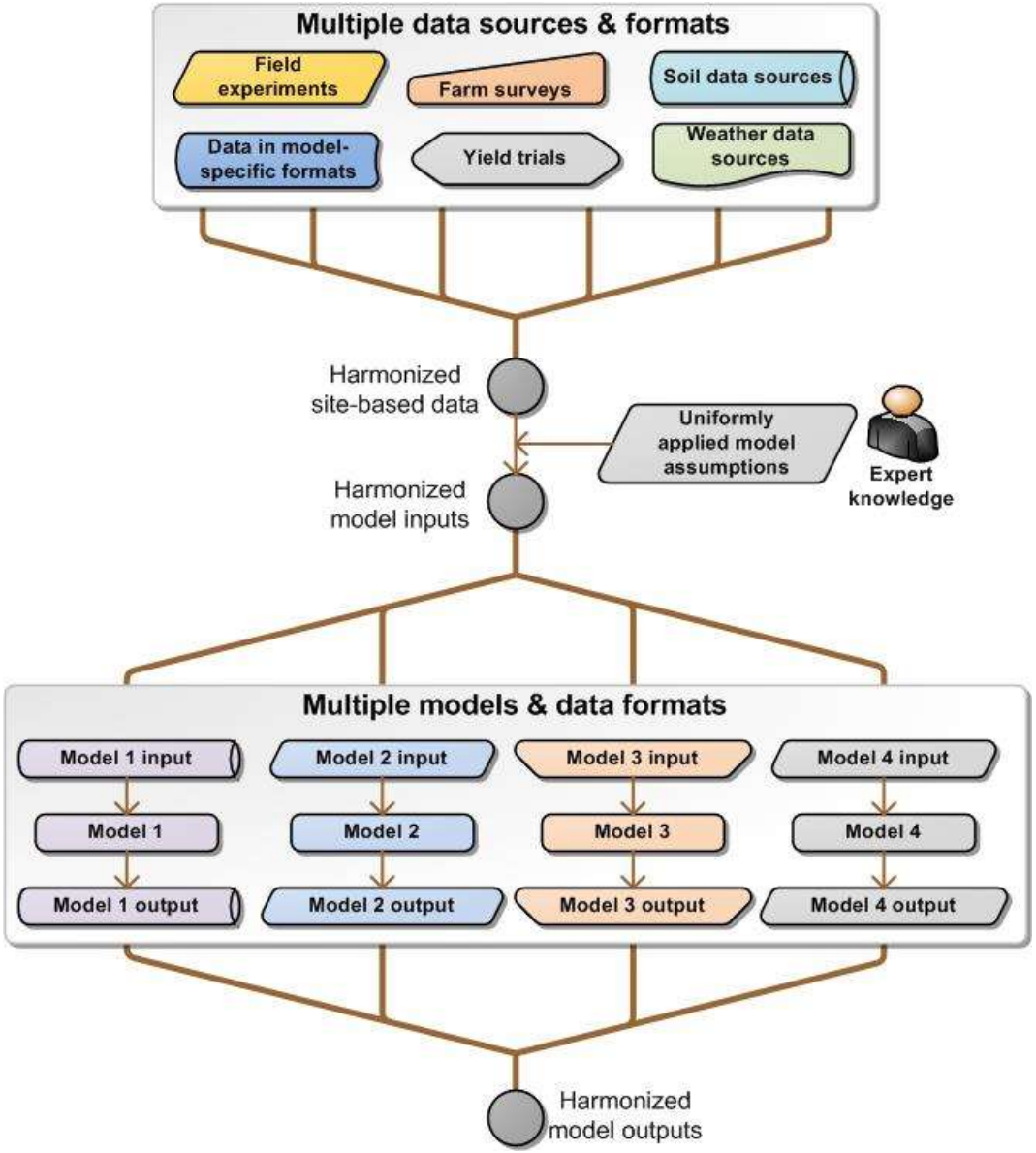
Allows data to be discovered and accessed.

Allows data to be searched in an automated way and connected with other databases through the semantic web.

Allows data to be transformed using existing AgMIP translators.

Allows fast, complex searching without accessing and interpreting the data directly.

Data interoperability tools



Designed to overcome

- Multiple inconsistent data sources
- Incomplete information to parameterize multiple crop models
- Multiple input and output formats for crop models

GARDIAN

HOME SEARCH ANALYTICS

DATA SEARCH

Search for

All ■ Open □ Restricted □

And ■ Or □

SEARCH

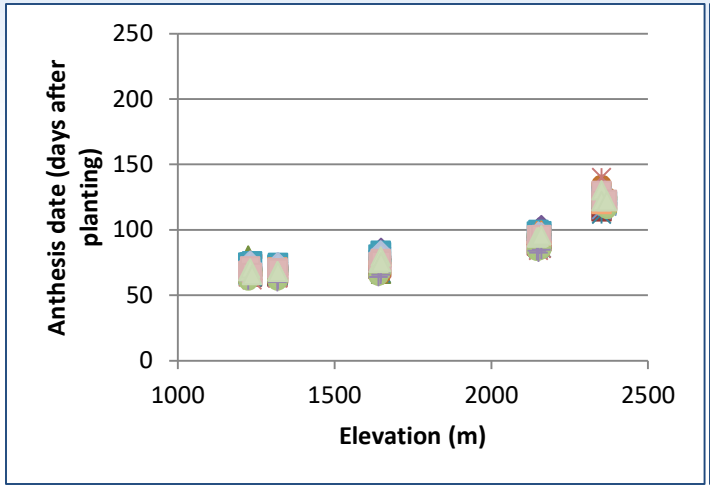
Resources What If

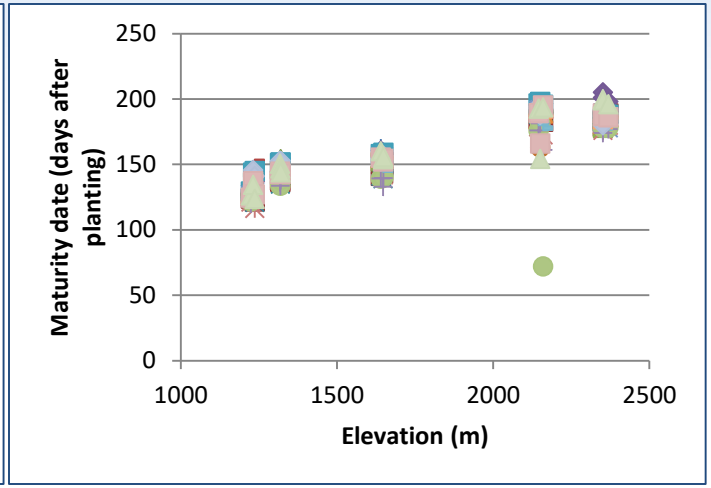
	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1	Temperature Average Maximum = AVMAXTemp	Maize variety name = MVnam	Planting date = PLNdat	Replicate = EXPrep	Plot number = PLOTnumb	Plot size (sq. m)	Emergence date = EMdat	Tasseling_silking_date=flowering_date = FLWdat	Pysiologic_maturity_dat = MDAT/date	Harvest_date = HDATE	Fresh biomass wt /plot (kg)	Dry biomass wt /plot (kg)	Grain yield /plot (kg)	Grain yield /ha (kg)	Grain yield t/ha = GRNyld	Harvest index (%)	Calc HW
2	32.5	BH660 (hybrid)	2016-05-23	1	1	14.4	2016-05-30	2016-08-03	2016-10-07	2016-11-28	38.92	33.78	13.56	9417.54	9.42	25.84	
3	32.5	PHB30-G19 (Shone - hybrid)	2016-05-23	1	2	14.4	2016-05-30	2016-07-26	2016-09-21	2016-11-28	26.89	19.32	8.93	6199.56	6.20	24.92	
4	32.5	BH543 (hybrid)	2016-05-23	1	3	14.4	2016-05-30	2016-08-03	2016-10-07	2016-11-28	21.19	16.35	8.25	5730.34	5.73	28.03	
5	32.5	BH540 (hybrid)	2016-05-23	1	4	14.4	2016-05-30	2016-07-28	2016-09-25	2016-11-28	22.17	16.90	7.92	5502.68	5.50	26.33	
6	32.5	P3812W (Limu- hybrid)	2016-05-23	1	5	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	16.94	14.02	4.91	3407.84	3.41	22.47	
7	32.5	AMH851(Jibat - hybrid)	2016-05-23	1	6	14.4	2016-05-30	2016-07-28	2016-09-25	2016-11-28	8.81	7.24	3.14	2179.45	2.18	26.27	
8	32.5	AMH760Q (hybrid)	2016-05-23	1	7	14.4	2016-05-30	2016-08-01	2016-10-03	2016-11-28	16.71	14.41	6.57	4563.64	4.56	28.23	
9	32.5	BHQPY545 (hybrid)	2016-05-23	1	8	14.4	2016-05-30	2016-07-28	2016-09-25	2016-11-28	25.01	19.30	9.13	6337.45	6.34	26.73	
10	32.5	MH138Q (hybrid)	2016-05-23	1	9	14.4	2016-05-30	2016-07-29	2016-09-27	2016-11-28	8.35	6.91	3.20	2225.17	2.23	27.73	
11	32.5	AMH853	2016-05-23	1	10	14.4	2016-05-30	2016-07-26	2016-09-21	2016-11-28	18.17	15.66	6.95	4824.75	4.82	27.66	
12	32.5	BH547 (hybrid)	2016-05-23	1	11	14.4	2016-05-30	2016-07-28	2016-09-25	2016-11-28	20.63	15.72	8.02	5566.54	5.57	27.98	
13	32.5	Melkasa-2 (OPV)	2016-05-23	1	12	14.4	2016-05-30	2016-07-24	2016-09-17	2016-11-28	14.00	10.29	5.68	3946.08	3.95	28.87	
14	32.5	BH546 (hybrid)	2016-05-23	1	13	14.4	2016-05-30	2016-08-29	2016-11-28	2016-11-28	16.03	13.54	5.64	3918.92	3.92	26.04	
15	32.5	MH140 (hybrid)	2016-05-23	1	14	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	15.54	14.49	6.14	4260.95	4.26	28.31	
16	32.5	AMH 854	2016-05-23	1	15	14.4	2016-05-30	2016-07-25	2016-09-19	2016-11-28	19.12	17.07	8.15	5663.08	5.66	29.90	
17	32.5	Melkasa-6Q * (OPV)	2016-05-23	1	16	14.4	2016-05-30	2016-07-23	2016-09-15	2016-11-28	7.34	6.04	3.09	2146.46	2.15	29.62	
18	32.5	Morka (OPV)	2016-05-23	1	17	14.4	2016-05-30	2016-08-05	2016-10-11	2016-11-28							
19	32.5	BH661 (hybrid)	2016-05-23	1	18	14.4	2016-05-30	2016-08-12	2016-10-25	2016-11-28							
20	32.5	Melkasa-4 (OPV)	2016-05-23	1	19	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28							
21	32.5	Gibe-2 (OPV)	2016-05-23	1	20	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28							
22	32.5	BH660 (hybrid)	2016-05-23	2	21	14.4	2016-05-30	2016-08-01	2016-10-03	2016-11-28	36.03	31.50	12.92	8974.36	8.97	26.40	
23	32.5	P3812W (Limu- hybrid)	2016-05-23	2	22	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	22.23	18.65	9.10	6316.62	6.32	29.03	
24	32.5	PHB30-G19 (Shone - hybrid)	2016-05-23	2	23	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	20.36	17.31	5.99	4160.82	4.16	22.74	
25	32.5	BH543 (hybrid)	2016-05-23	2	24	14.4	2016-05-30	2016-08-05	2016-10-11	2016-11-28	12.54	10.88	4.11	2851.85	2.85	24.67	
26	32.5	BH540 (hybrid)	2016-05-23	2	25	14.4	2016-05-30	2016-07-29	2016-09-27	2016-11-28	17.89	14.61	7.50	5210.60	5.21	29.55	
27	32.5	BHQPY545 (hybrid)	2016-05-23	2	26	14.4	2016-05-30	2016-08-01	2016-10-03	2016-11-28	19.99	16.55	6.70	4650.48	4.65	25.09	
28	32.5	AMH851(Jibat - hybrid)	2016-05-23	2	27	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	14.05	12.03	5.13	3561.64	3.56	26.74	
29	32.5	AMH760Q (hybrid)	2016-05-23	2	28	14.4	2016-05-30	2016-08-03	2016-10-07	2016-11-28	19.47	16.44	5.74	3985.60	3.99	22.77	
30	32.5	AMH 853	2016-05-23	2	29	14.4	2016-05-30	2016-07-27	2016-09-23	2016-11-28	13.98	9.34	3.95	2745.10	2.75	22.05	
31	32.5	MH140 (hybrid)	2016-05-23	2	30	14.4	2016-05-30	2016-07-28	2016-09-25	2016-11-28	31.01	13.44	2.07	5531.34	5.53	26.66	

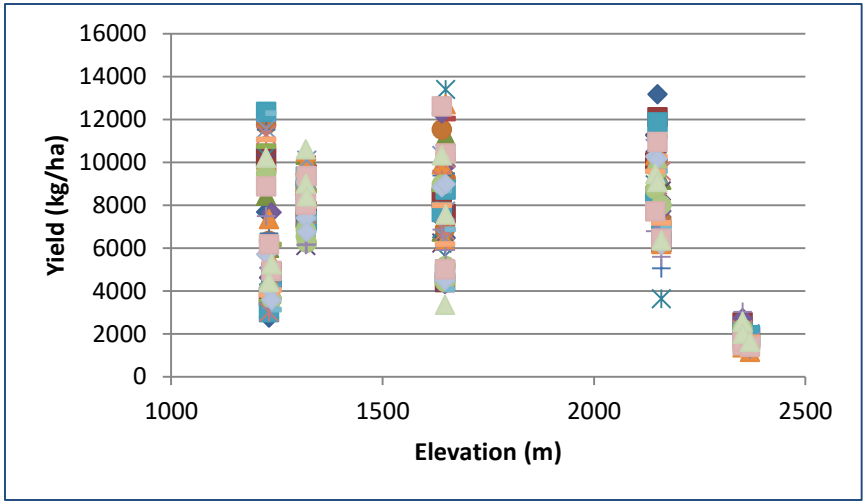
Metadata Protocol GxE Layout Variables_modified Raw Data Corrected Data Sidecar #1 Sidecar #1 (2) translation basis Pivot Summary Temperature_Data

Phenology and yield vs elevation

- AMH 853
- AMH 854
- AMH760Q (hybrid)
- AMH851(Jibat - hybrid)
- AMH853
- BH540 (hybrid)
- BH543 (hybrid)
- BH546 (hybrid)
- BH547 (hybrid)
- BH660 (hybrid)
- BH661 (hybrid)
- BHQPY545 (hybrid)
- Gibe-2 (OPV)
- Melkasa-2 (OPV)
- Melkasa-4 (OPV)
- Melkasa-6Q * (OPV)
- MH138Q (hybrid)
- MH140 (hybrid)
- Morka (OPV)
- P3812W (Limu- hybrid)
- PHB30-G19 (Shone - hybrid)

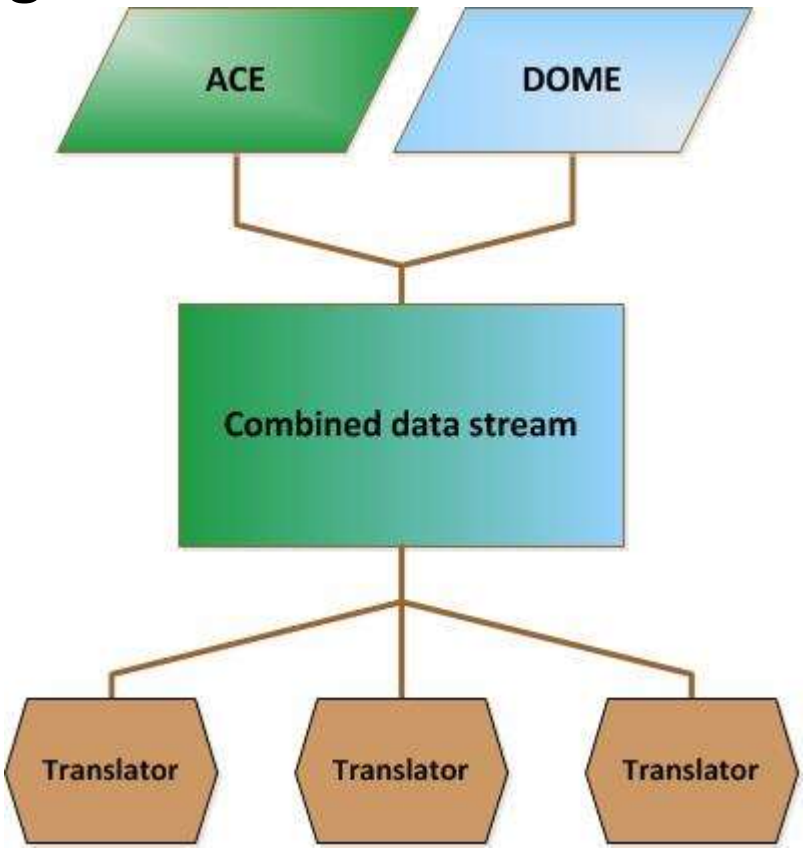




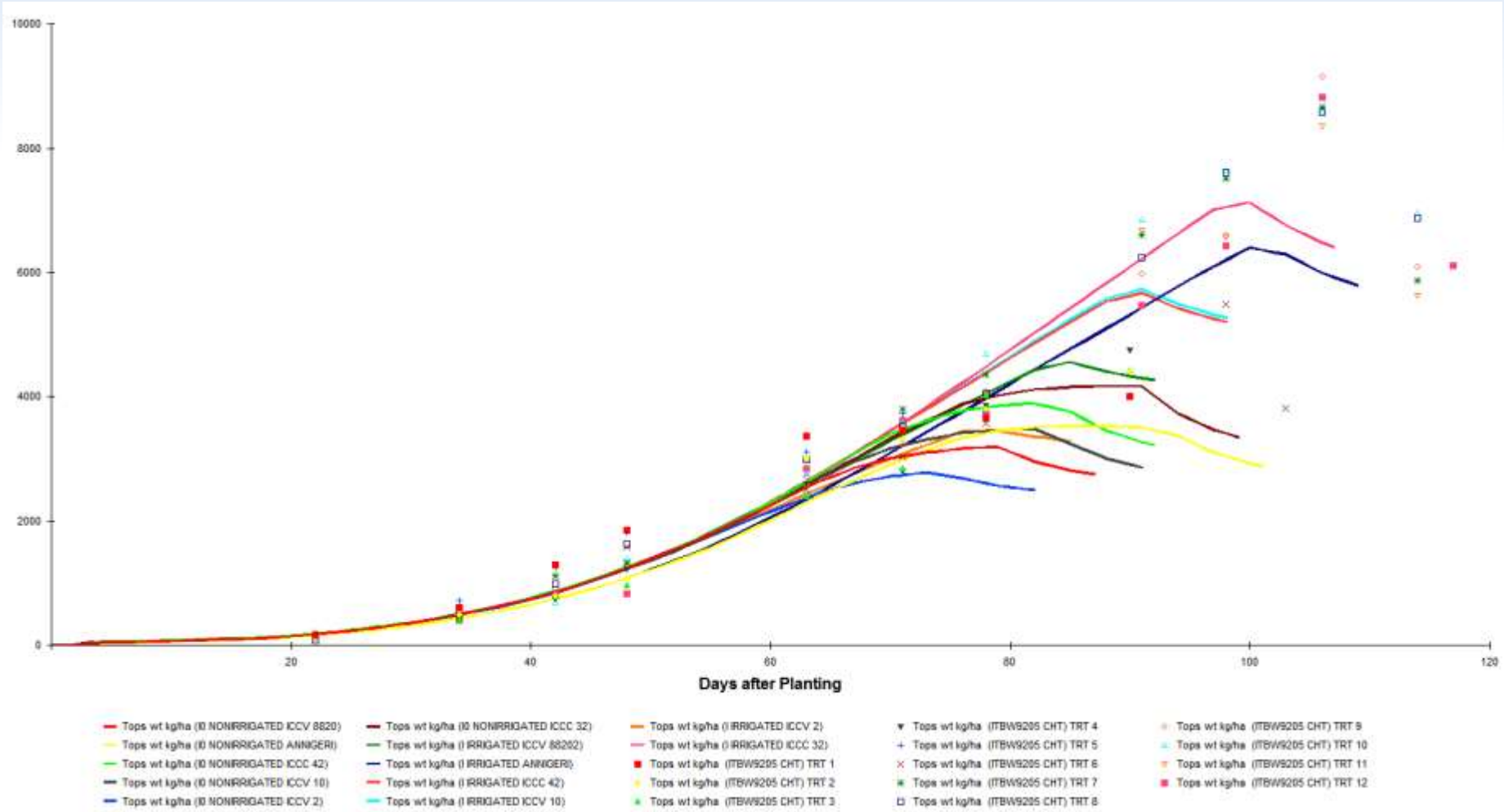


DOME data

- supplemental information needed for crop models, but not supplied by data
  - initial soil moisture and nitrogen content
  - planting depth
  - fertilizer type
  - …



DSSAT simulation



```
[
  {
    "data": "GPS Coordinate Latitude",
    "term": "http://purl.obolibrary.org/obo/OBI_0001620",
    "unit": "http://purl.obolibrary.org/obo/UO_0000185"
  },
  {
    "data": "GPS Coordinate Longitude",
    "term": "http://purl.obolibrary.org/obo/OBI_0001621",
    "unit": "http://purl.obolibrary.org/obo/UO_0000185"
  },
  {
    "data": "Planting date = PLNdat",
    "term": "http://purl.obolibrary.org/obo/AGRO_00000231",
    "unit": "http://purl.obolibrary.org/obo/AGRO_00000417"
  },
  {
    "data": "Harvest_date = HDATE",
    "term": "http://purl.obolibrary.org/obo/AGRO_00000025",
    "unit": "http://purl.obolibrary.org/obo/AGRO_00000417"
  },
  {
    "data": "Grain yield /ha (kg)",
    "term": "http://purl.obolibrary.org/obo/TO_0000371",
    "unit": "http://purl.obolibrary.org/obo/UO_0000283"
  },
  {
    "data": "Site no. =EXPsit",
    "term": "http://purl.obolibrary.org/obo/NCIT_C90472"
  },
  {
    "data": "Plot size (sq. m)",
    "term": "http://purl.obolibrary.org/obo/AGRO_00000302",
    "unit": "http://purl.obolibrary.org/obo/UO_0000080"
  },
  {
    "data": "Harvest index (%)",
    "term": "http://purl.obolibrary.org/obo/TO_0000128",
    "unit": "http://purl.obolibrary.org/obo/UO_0000187"
  }
]
```

Annotation – Sidecar file 1

Annotation – Sidecar file 2


```
{
  "info": {
    "datatype": "excel",
    "default_sheet": "Corrected Data",
    "translator_email": "cvillalobos@ufl.edu",
    "source": "https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/11011"
  },
  "translation_map": [
    {
      "icasa": "crid",
      "provider": "user",
      "fixed_value": "MAZ"
    },
    {
      "icasa": "fl_lat",
      "provider": "data",
      "column": "GPS Coordinate Latitude",
      "unit": "UO_0000185"
    },
    {
      "icasa": "fl_long",
      "provider": "data",
      "column": "GPS Coordinate Longitude",
      "unit": "UO_0000185"
    },
    {
      "icasa": "pdate",
      "provider": "data",
      "column": "Planting date = PLNdat",
      "unit": "AGRO_00000417",
      "format": "excel"
    },
    {
      "icasa": "hdat",
      "provider": "data",
      "column": "Harvest_date = HDATE",
      "unit": "AGRO_00000417",
      "format": "mm/dd/yy"
    },
    {
      "icasa": "hwam",
      "provider": "data",
      "column": "Grain yield /ha (kg)",
      "unit": "UO_0000283"
    },
    {
      "icasa": "plpop",
      "provider": "calculated",
      "formula": "10^4/(plrs * plsp)",
      "unit": "--NA: Count per square meter--"
    },
    {
      "icasa": "exname",
      "provider": "method",
      "method": {
```

Annotation – Sidecar file 3


```
{
  "crid": [
    "MAZ"
  ],
  "locations": [
    "8.967523167,37.85994389",
    "8.967607487,37.85981714",
    "8.967685209,37.85973552",
    .....
    "9.418214133,36.5396118",
    "9.418323846,36.5398268",
    "9.41844009,36.53997379"
  ],
  "pdate": [
    "2016-05-13",
    "2016-05-20",
    "2016-05-23",
    .....
    "2016-06-07",
    "2016-06-12",
    "2016-06-15"
  ],
  "adat": [
    "2016-07-13",
    "2016-07-18",
    "2016-07-20",
    .....
    "2016-10-13",
    "2016-10-14",
    "2016-11-06"
  ],
  "mdat": [
    "2016-03-12",
    "2016-09-12",
    "2016-09-15",
    .....
    "2017-01-02",
    "2017-01-03",
    "2017-01-04"
  ],
  "hdate": [
    "2016-11-28",
    "2016-11-29",
    "2016-12-09",
    .....
    "2017-01-03",
    "2017-01-10",
```