

# **BETYdb data collection:**

**Biofuel Ecophysiological Traits and Yields database**

David LeBauer

Energy Biosciences Institute

July 23, 2014

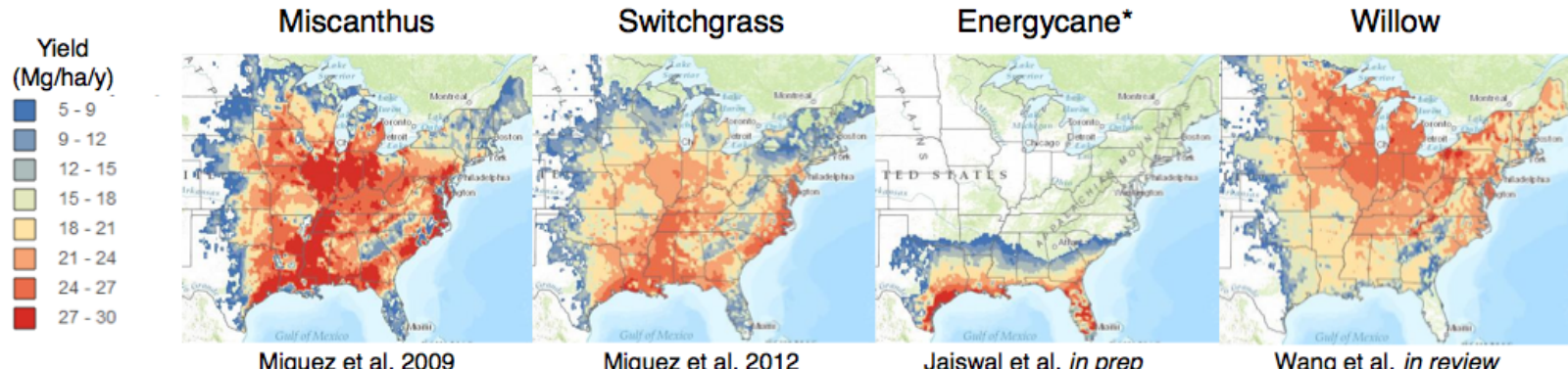
# Biofuel Feedstock production

What species?

Where can they be grown?

How much can they produce?

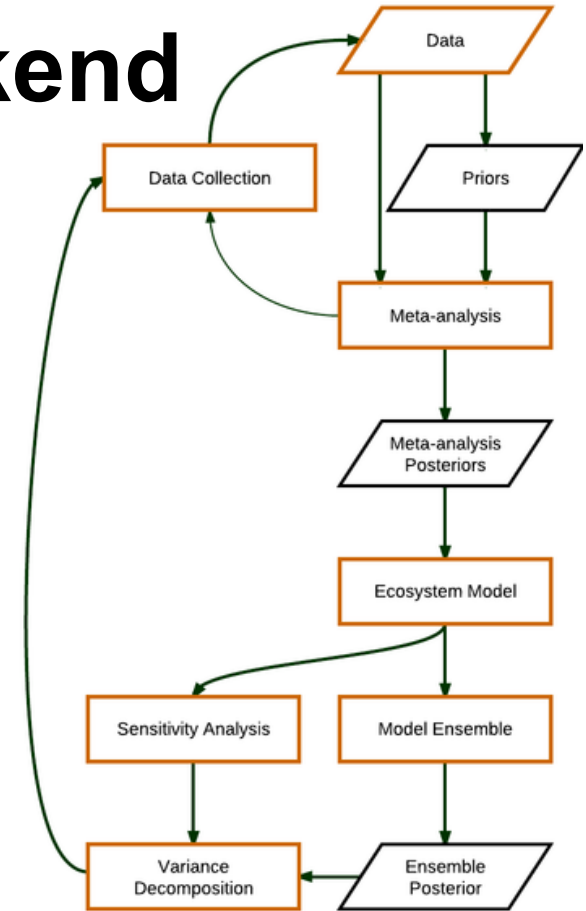
What are ecological consequences?



# BETYdb: PEcAn's informatics backend

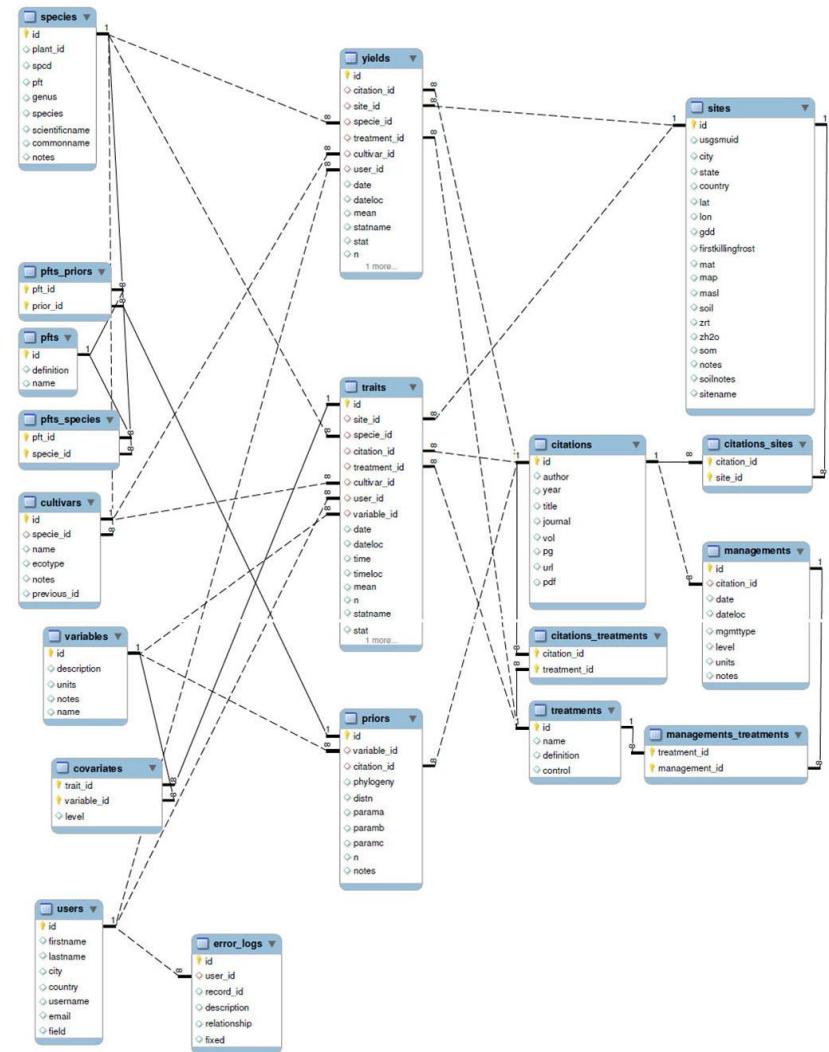
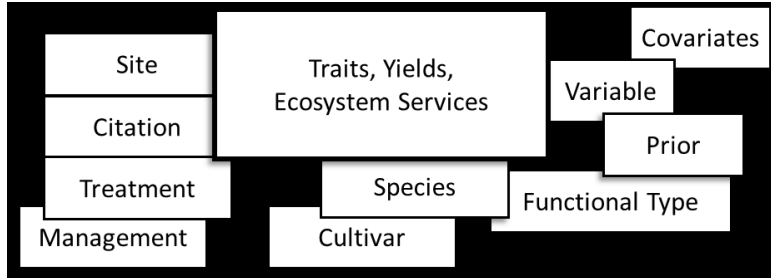
## Contents:

- parameterization data (traits)
- validation data (yields, ecosystem functions)
- drivers (meterology, soil)
- results (model output, parameter estimates)
- provenance



# Schema

Tables used to store trait, yield data



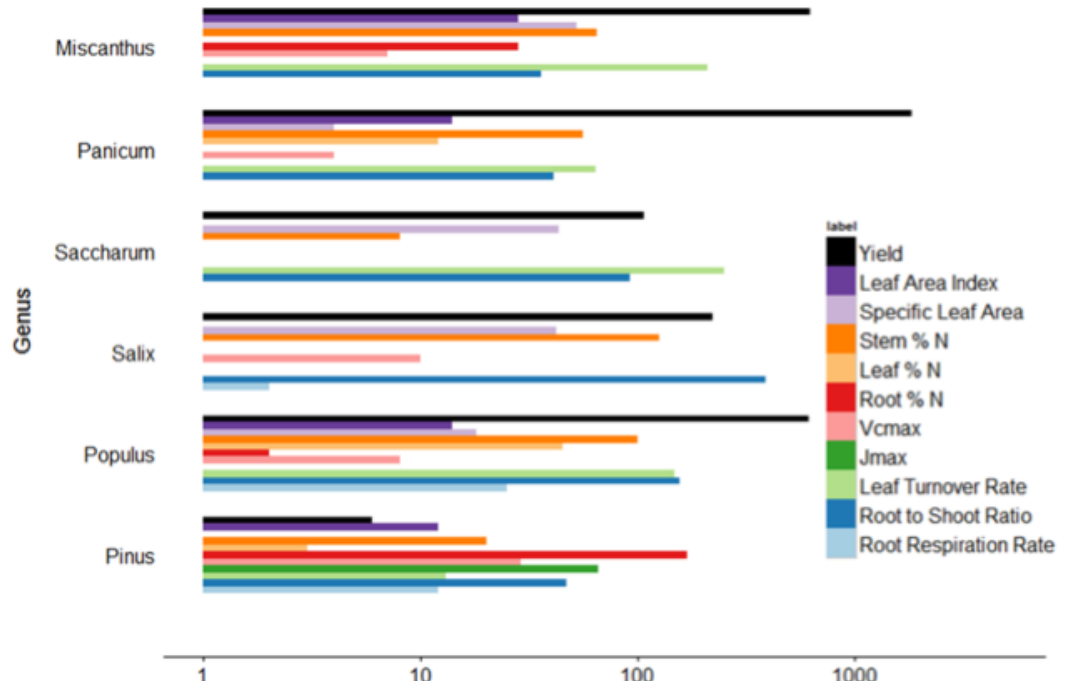
# Current Trait and Yield data

30k plant traits ( $\sim\frac{1}{3}$  biofuels)

8k biofuel yields

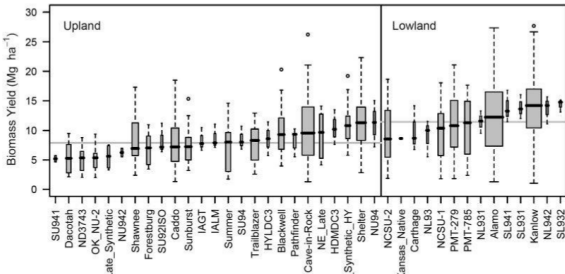
\* “Published”

\* “Raw”

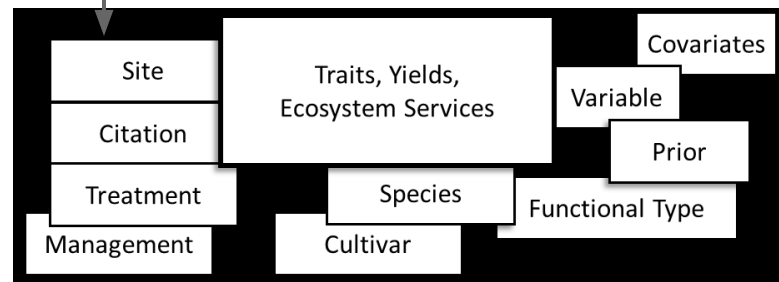


# Published data

- \* primary focus to date
- \* [10 page documentation](#)
- \* Separate scientific understanding from data entry
- \* Difficult even for experts to interpret
- \* May not be possible to automate



Indicators	$N_0$	$N_{60}$	$N_{120}$	LSD <sub>05</sub>	$N_0$	$N_{60}$	$N_{120}$	LSD <sub>05</sub>
Second year of growth								
Annual biomass								
t ha <sup>-1</sup>	15.8	20.0	24.7*	5.63	4.62	6.55*	6.82*	1.85
%	100	127	157	40.5	100	142	148	42.9
Biomass weight per plant								
kg	1.66	2.26*	2.53*	0.627	0.49	0.74*	0.70*	0.208
Third year of growth								
Annual biomass								
t ha <sup>-1</sup>	27.0	28.5	29.7	5.31	10.5	10.7	11.5	2.47
%	100	105.6	110.1	18.68	100	102.3	110.2	22.61
Biomass weight per plant								
kg	2.05	2.18	2.25	0.396	0.79	0.81	0.87	0.187



# Raw Data

No standard format.

Lots of it

More informative, valuable

BETYdb as archive for EBI data

	A	B	C	D	E	F	G	H	I	J	K
1	Site	Date	Species	Genotype	Type	Block	IRGA	Curve		Topt	PAR
2	NY	5/24/13	Willow	SX61	Sun	1	Dietze	Temp		27.25	1500
3	NY	5/25/13	Willow	SX61	Sun	1	Dietze	Temp		Bad Curve	1500
4	NY	5/24/13	Willow	SX61	Sun	2	Dietze	Temp		Bad Curve	1500
5	NY	5/25/13	Willow	SX61	Sun	2	Dietze	Temp		26.332	1500
6	NY	5/24/13	Willow	SX61	Sun	3	Dietze	Temp		Bad Curve	1500
7	NY	5/24/13	Willow	FC	Sun	1	USDA	Temp		Bad Curve	1500
8	NY	5/25/13	Willow	FC	Sun	1	USDA	Temp		Bad Curve	1500
9	NY	5/24/13	Willow	FC	Sun	2	USDA	Temp		23.6	1500
10	NY	5/25/13	Willow	FC	Sun	2	USDA	Temp		25.9	1500
11	NY	5/24/13	Willow	FC	Sun	3	USDA	Temp		Bad Curve	1500

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Miscanthus	MCB	SD	MNB	SD	MCM	SD	MNM	SD	MCT	SD	MNT	SD
2	Jun	19.73	7.48	21.28	2.38					33.02	6.44	40.35	7.26
3	Jul	11.45	2.96	18.88	13.13	15.99	0.70	9.69	1.74	27.83	3.46	31.45	3.87
4	Aug	10.37	3.68	13.00	2.14	21.25	7.18	15.81	3.76	27.61	8.68	22.01	7.21
5	Sep	12.37	5.06	12.24	4.15	15.85	2.07	22.80	3.19	23.48	3.63	21.93	3.58
6	Oct	23.16	10.35	20.27	11.75	17.06	2.08	18.55	9.11	31.51	4.08	30.41	2.19
7													
8													
9													
10	Switchgrass	CB	SD	NB	SD	CM	SD	NM	SD	CT	SD	NT	SD
11	Jun									25.62	4.38	29.43	5.30
12	Jul	10.53	0.89	11.44	2.42					23.05	2.09	20.41	4.14
13	Aug	13.37	2.97	11.96	3.94					12.61	3.41	23.91	5.92
14	Sep	10.71	1.01	11.97	3.37					12.48	4.95	8.83	1.51

# Importing Raw data

- \* Trying to develop a template
  - \* easy for researchers to use
  - \* can be automatically parsed
- \* Use web-forms for meta-data (site, citation, treatments)
- \* Substituting text fields for primary keys in raw data files
- \* Identifying related records using both wide and long formats (entities)

	A	B	C	D	E	F	G	H	I
1		species	cultivar	site	treatment	date	yield	SE	n
2	1	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	0.2	-9999	4
3	2	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	0.3	0.3	4
4	3	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	0.4	-9999	4
5	4	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	0.5	-9999	4
6	5	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	3.25	1.25	4
7	6	Miscanthus x giganteus		Southeast Georg	Observational Ge	12/17/2010	4.4833	0.2167	4



# Opportunities to use Brown Dog

Raw data from EBI, collaborators, repositories

- \* Need to convert from a set of 'native' formats to betydb schema
  - \* no conventions exist, need to find a middle ground
  - \* try-db, sungrant, bioenergy kdf, DataOne
  - \* describe BETYdb with EML to facilitate interoperability

Other Opportunities:

- \* canopy geometry from images
- \* homogenization of soil databases
- \* ability to estimate uncertainty

# More Info

website: [betydb.org](http://betydb.org)

code: [github.com/PecanProject/bety](https://github.com/PecanProject/bety)

funding: Energy Biosciences Institute, NSF

contact: [dlebauer@illinois.edu](mailto:dlebauer@illinois.edu)