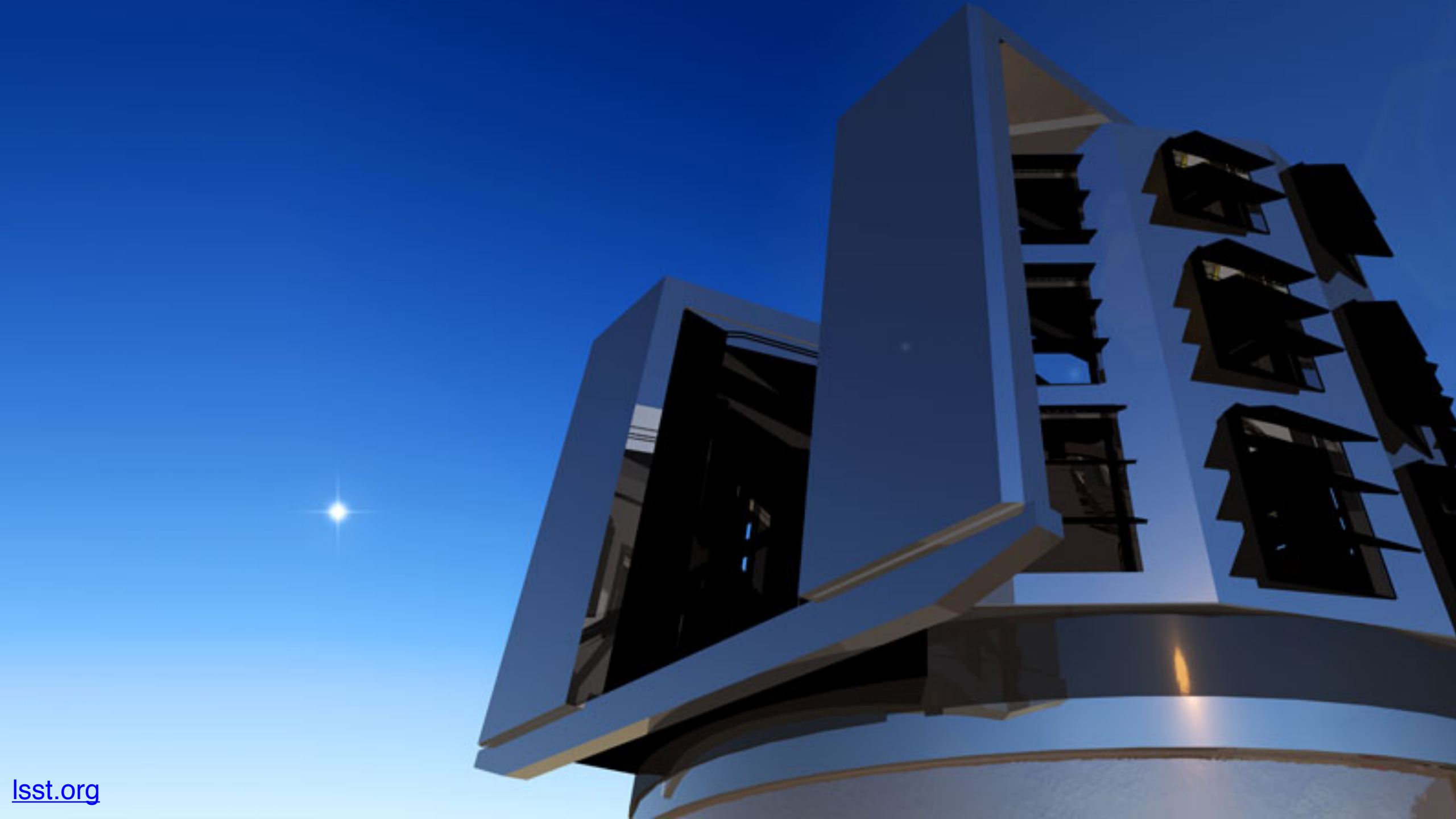
TERRA Reference Data and Computing Platform

David LeBauer, University of Illinois and NCSA

5th National Data Service Consortium Workshop

Chapel Hill, NC

April 5, 2016



lsst.org

https://youtu.be/Pp6IdkPtFC8

ARPA-E TERRA Objective: Faster Plant Breeding with Sensors

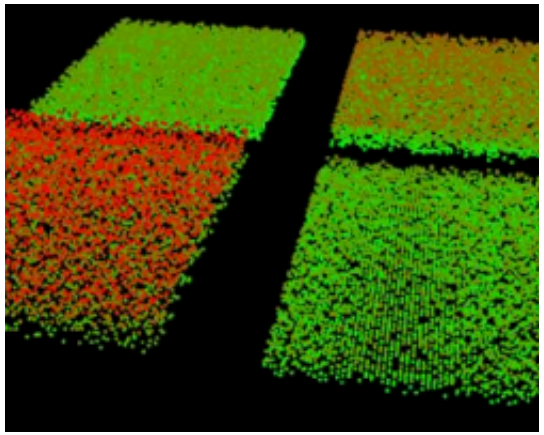
Remote, Robotic Sensing for Plant Breeding

- "High Throughput Phenotyping"
- Screen more plants & more traits

Approach

- Deploy Modern Sensor Suite
- Capture Data with Unprecedented Resolution (~daily @mm)
- Open Data, Accessible Computing

http://bulletin.ipm.illinois.edu/print.php?id=513



ARPA-E TERRA: Common Dataset for Six Projects

Public Reference Data Team

- Donald Danforth Plant Science Center — Discovery | Community | Impact
- Washington University in St. Louis
- USDA ARS
- NCSA
- Illinois
- The University of Arizona
- HudsonAlpha Institute for Biotechnology
- Kansas State University
- JGI
- Texas A&M University
- LemnaTec
- Clemson University

Full System Teams

- Purdue University — College of Agriculture, College of Biological Engineering, College of Aeronautics
- IBM Research
- CSIRO — Australia Plant Phenomics: Commonwealth Scientific and Industrial Research Organization
- Clemson University — College of Agriculture, Forestry, and Life Sciences
- Carnegie Mellon Robotics Institute
- Near Earth Autonomy
- Donald Danforth Plant Science Center

"Component" Focus Systems Teams

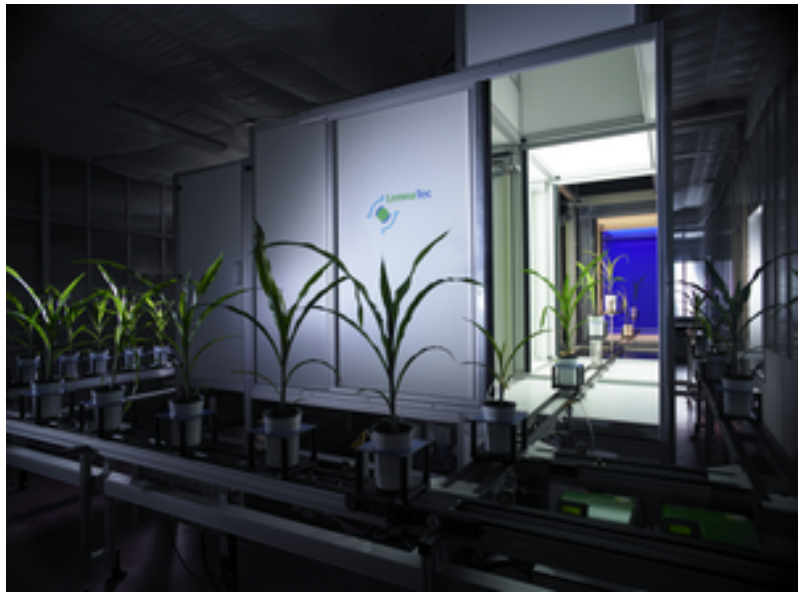
- Texas A&M AgriLife Research
- National Robotics Engineering Center (NREC)
- ceres
- Pacific Northwest National Laboratory
- Chromatin
- Berkeley Lab — Lawrence Berkeley National Laboratory
- Blue River Technologies
- University of California Agriculture and Natural Resources
- Illinois — University of Illinois at Urbana-Champaign
- Cornell University — College of Agriculture and Life Sciences
- Berkeley EECS — Electrical Engineering & Computer Sciences
- USDA ARS

Domains:

- Plant Breeding
- Ecophysiology
- Genomics
- Computer Vision
- Robotics
- Statistics
- Simulation

ARPA-E TERRA

TERRA Reference Data Sources



Lemnatec Scanalyzer
Danforth, St. Louis

1TB / month
20 TB total

Lemnatec Field System
USDA ALRC, Maricopa, AZ

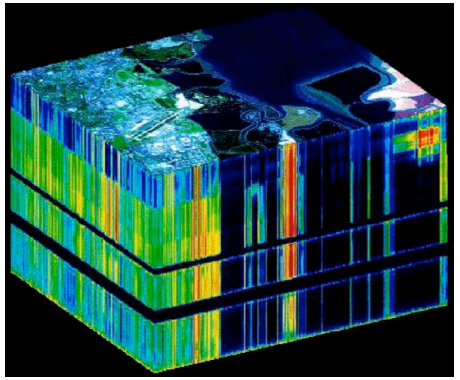
~1-5 TB/d
1-10 PB total

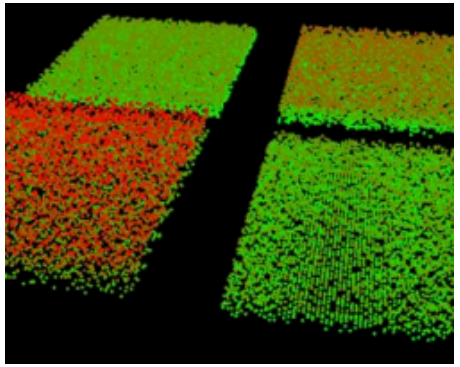
Tractor and UAV
Kansas State

1 TB / month
100 TB total

Also, contributions from other teams, community

Field Sensors





VNIR Imaging Spectrometer 380-1000nm

SWIR Imaging Spectrometer 900-2500 nm

IR Temperature Sensor

NDVI (1 down, 1 up) 650, 800 nm

PRI Sensor 531, 570 nm

PAR Sensor 410-655 nm

Color Sensor 410-655 nm

3D Scanners: 2 Side View, 1 Down

RGB: 2 Side View, 1 Down (1)

Active Reflectance 670, 730, 780 nm

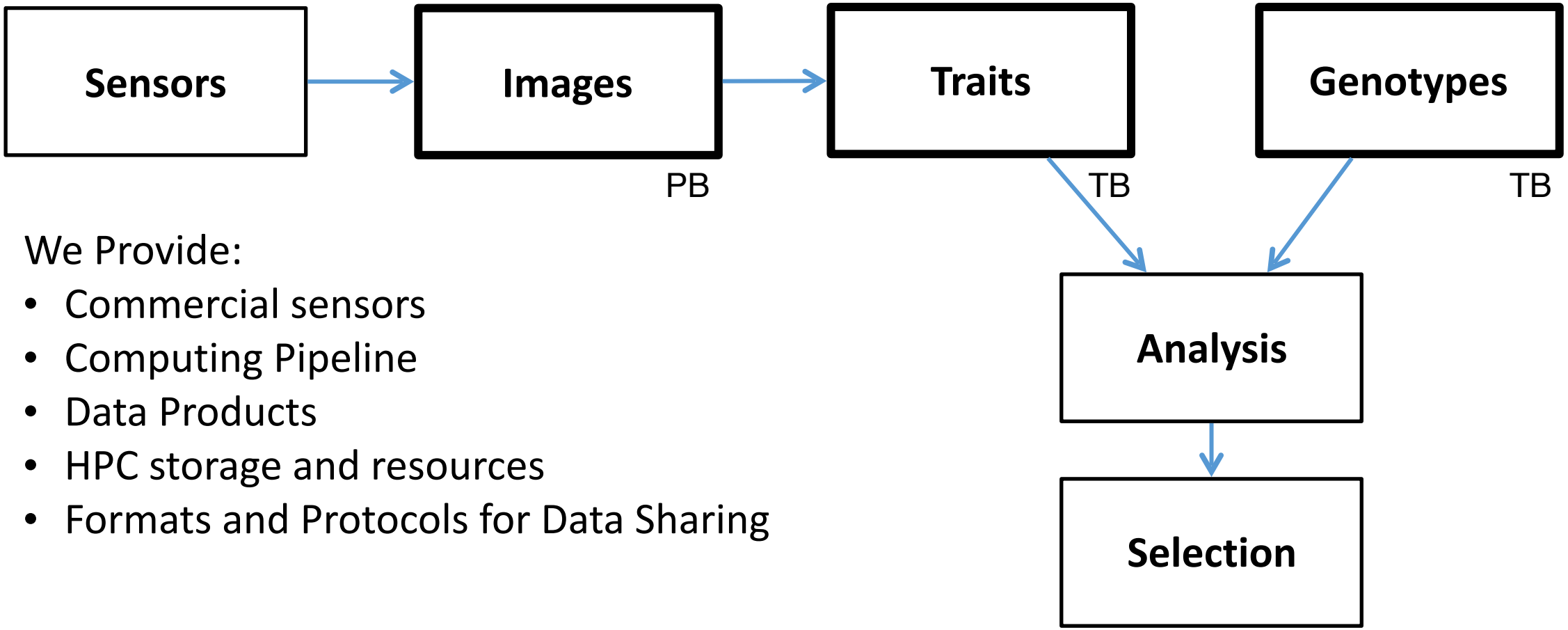
Fluorescence

Environmental:

wind, temperature, humidity, pressure, light, rain, CO2

terraref.ncsa.illinois.edu/articles/lemnatec-scanalyzer-field-sensors/

TERRA REF Pipeline



We Provide:

- Commercial sensors
- Computing Pipeline
- Data Products
- HPC storage and resources
- Formats and Protocols for Data Sharing

TERRA REF Data Products

Sensor

- Images
- Rasters, clouds, surfaces
- Environmental

Traits

- Biomass
- Height
- Stem diameter, no.
- Chemistry
- Photosynthesis
- Stress Tolerance

Genomics

- 1000s Sorghum Lines
- Raw Sequence Data
- Aligned Reads
- SNPs

What is Available

Sensors → **Images** → **Traits**

Genotypes

Analysis

Selection

lemnatec.com

teamcssi.com

Not Scalable, Ad-hoc data formats, Proprietary

What Users Want

IDEs:

Emacs, Vi, Jupyter, Rstudio

Languages:

Python, Perl, C++, R, …

Software:

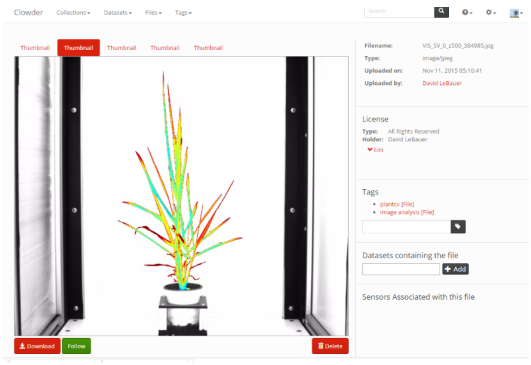
PlantCV, NCO, BWA, GTK

APIs & Standard Formats:

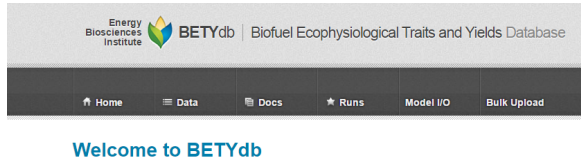
BETYdb, BRAPI, CF, ICASA

Web Services / Databases:

Clowder, BETYdb, BMS, CoGe, THREDDS

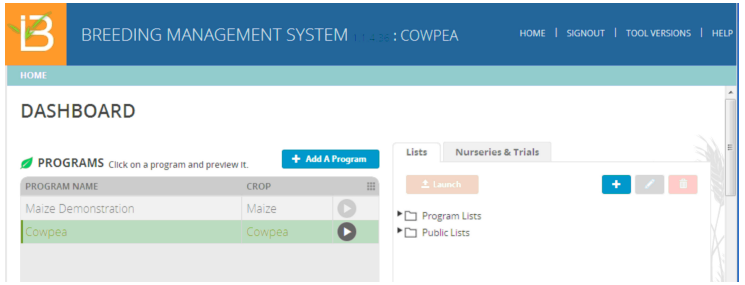


Clowder



BETYdb.org

CoGe



BMS

And more, based on need

Scalable, Modular, Interoperable, Open-source

Acknowledgements

Technical Team

Noah Fahlgren	Danforth Center
JD Maloney	NCSA / SET
Rob Kooper	NCSA / ISDA
Yan Liu	NCSA / CyberGIS
Max Burnette	NCSA / ISDA
Charlie Zender	UC Irvine
Scott Rohde	University of Illinois

Collaborators

Todd Mockler	Danforth Center
Noah Fahlgren	
Stephen Kresovich	Clemson University
Jeremy Schmutz	HudsonAlpha
Jesse Poland	Kansas State University
Geoff Morris	
William Rooney	Texas A&M University
Pedro Andrade-Sanchez	University of Arizona
Michael Ottman	
Jeff White	USDA-ARS
David LeBauer	University of Illinois
Robert Pless	Washington University
Roman Garnett	

Data Products & Standards Committee

David Lee	ARPA-E
David LeBauer	UIUC / NCSA
Paul Bartlett	Near Earth Autonomy
Jeff White	USDA ALARC
Melba Crawford	Purdue
Michael Gore	Cornell University
Matt Colgan	Blue River
Christer Janssen	PNNL
Barnabas Poczos	Carnegie Mellon
Alex Thomasson	Texas A&M University
Cheryl Porter	ICASA / AgMIP / USDA
Shawn Serbin	Brookhaven National Lab
Shelly Petroy	NEON
Carolyn J. Lawrence-Dill	Iowa State
Eric Lyons	University of Arizona, CyVerse

arpa·e CHANGING WHAT'S POSSIBLE

NCSA

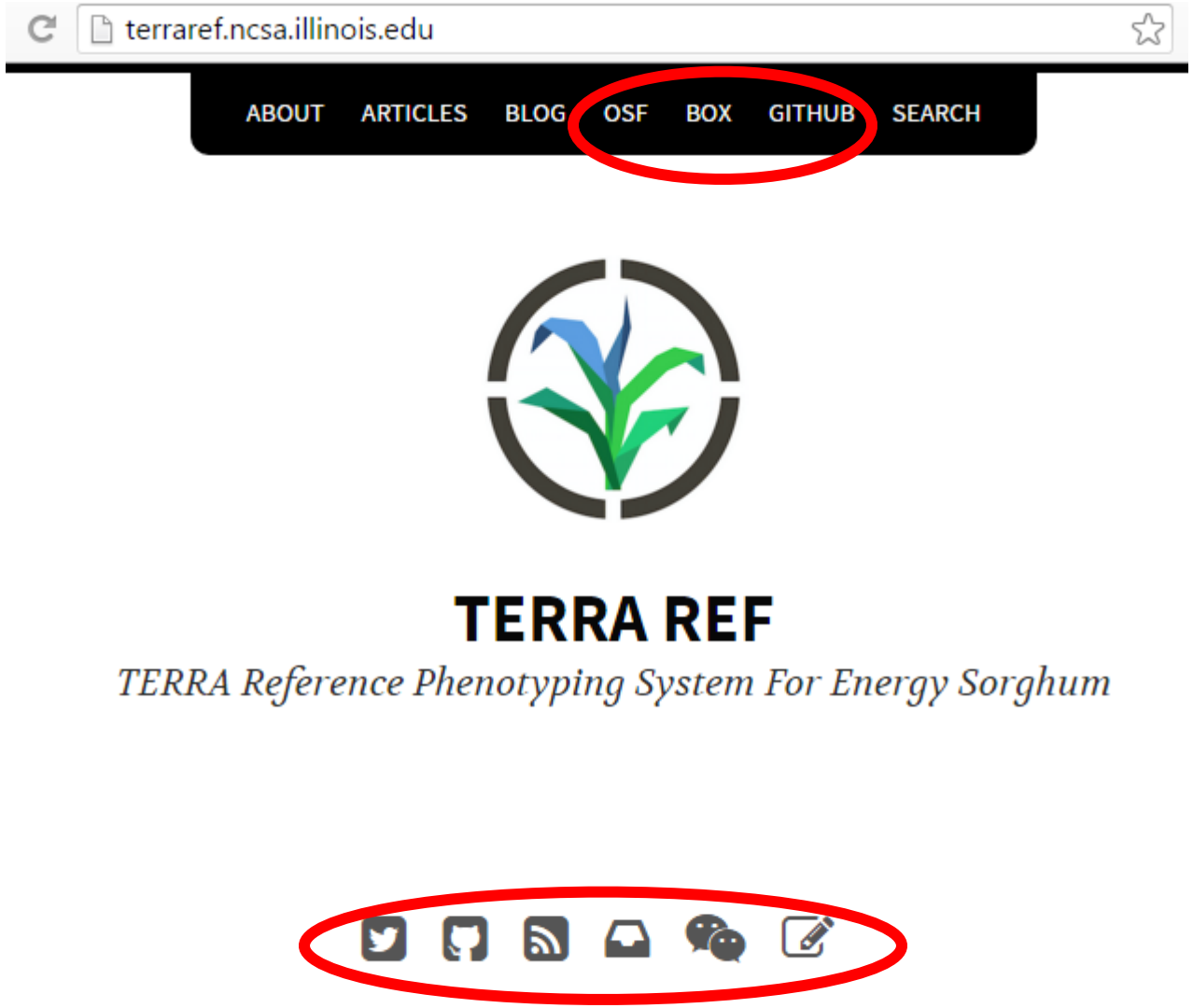
LemnaTec

Contact

UI / NCSA Lead: David LeBauer
dlebauer@illinois.edu

@terra_ref

terraref.ncsa.illinois.edu



Data Standards / Data Products Committee

Participants

- One from each project
- Six External Experts
- Community*

Responsibilities

- Define Data Products
- Identify & Link Standards
- Develop new formats

Activities

- Monthly Meetings
- Annual Workshops
- Revise, Request, Improve
- Training and Outreach

*Github Issues:
github.com/terraref/reference-data/issues

TERRA REF mandate

- High-resolution dataset from current technology
- Computing Pipeline
- Interactive Analysis and Development
- Deployable Technology
- Archiving, Sharing, and Publication of Reproducible Workflows
- Open [Science, Data, Software]
- Useable, Useful, Familiar: Scientists, Breeders, Precision Ag
- Modular, Extensible, Distributed
- Couple Existing Tools / Databases / APIs
- Scalable: From Mobile Devices to High Performance Computers

Computing and Storage Resources

- **Roger:** CyberGIS 1PB allocated, diverse hardware, GIS / big data optimized
- **Nebula:** NCSA Production Open Stack Server
- **Blue Waters**: +2.5 PB / y tape storage for TERRA

Standards Committee

Name	Institution
Coordinators	
David Lee	ARPA-E
David LeBauer	UIUC / NCSA
TERRA Project Representatives	
Paul Bartlett	Near Earth Autonomy
Jeff White	USDA ALARC
Melba Crawford	Purdue
Michael Gore	Cornell University
Matt Colgan	Blue River
Christer Janssen	PNNL
Barnabas Poczos	Carnegie Mellon
Alex Thomasson	Texas A&M University
External Advisors	
Cheryl Porter	ICASA / AgMIP / USDA
Shawn Serbin	Brookhaven National Lab
Shelly Petroy	NEON
Carolyn J. Lawrence-Dill	Iowa State
Eric Lyons	University of Arizona, iPlant

TERRA
REF