

Containerized development environments for large sensor data

David LeBauer

National Center for Supercomputing Applications
Agricultural and Biological Engineering, University of Illinois



TERRAPHENOTYPING
REFERENCE PLATFORM

High Throughput Phenotyping

Phenotyping: Measuring Plant Properties

High Throughput: automating measurements with sensors, robots, drones, etc

Approach: Open Access Reference Data

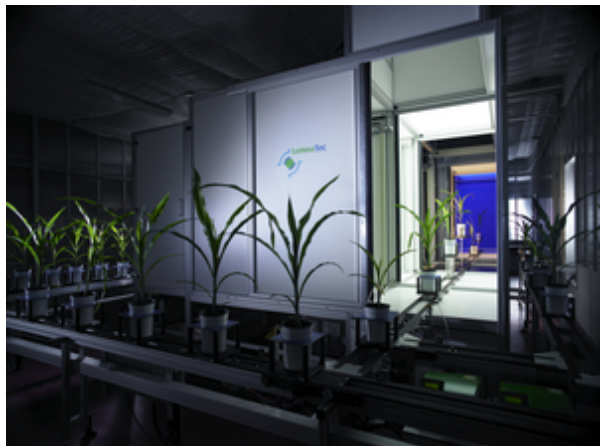
Modern Sensor Suite

Unprecedented Resolution (~daily @mm)

Open Data, Accessible Computing



TERRA Reference Data Sources



Lemnatec Scanalyzer
Danforth, St. Louis

1TB / month
5 TB total



Lemnatec Field System
USDA ALRC, Maricopa, AZ

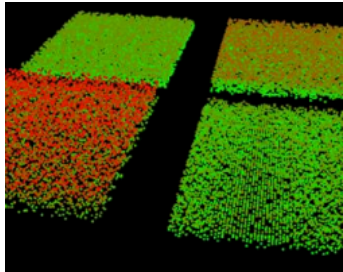
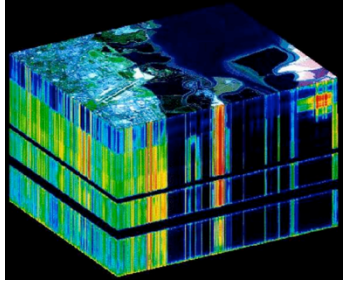
~1-5 TB/d
1-10 PB
total



Tractor and UAV
Kansas State

1 TB / month
100 TB total

Lemnatec 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

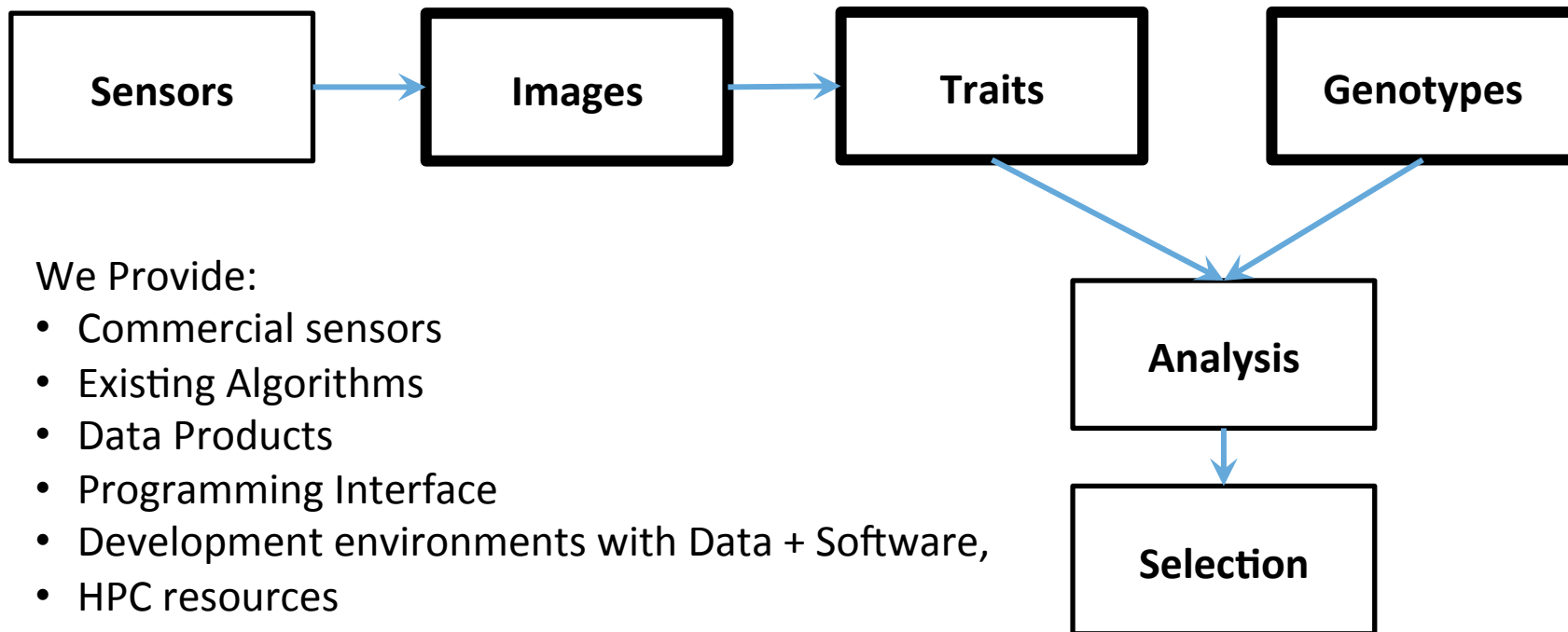
RGB: 2 Side View, 1 Down (1)

Active Reflectance 670, 730, 780 nm

Fluorescence

Environmental: **wind, temperature, humidity, pressure, light, rain, CO2**

Computing Pipeline and Data Products



TERRA Reference Design Principles

- Open [Science, Data, Software]
- Useable, Useful, Familiar: Scientists, Breeders, Precision Ag
- Modular, Extensible, Distributed
- Reproducible, Collaborative
- Scalable: From Mobile Devices to High Performance Computers

Technical Challenges

- Individual files commonly 10-100s GB; a single day ~TB
- Data and software is heterogeneous
- Large training and collaboration component
- Reproducibility

How Containers Help

Development

- NDS Workbench provides a primary portal to large data
- Researchers can prototype algorithms before deployment

Training

- Workshops & Hackathons
- Tutorials: interactive documentation

Version Control and Archiving

- Publication and reproducible science
- Dataset provenance

Development: Find and Use Data

Maricopa Agricultural Center Field Scanner

Delete Create Dataset Create Collection

Collections

Most recent collections:

All Collections

Lemnatec Field Scanner Calibration Data

1 0

2016-02-16

All datasets on this date.

137 0

2016-02-18

All datasets on this date.

3 0

2016-02-19_

All datasets on this date.

1 0


2016-05-02

7255 0

flirIrCamera

All datasets from this sensor.

1 0



Maricopa Field Data and Notes

Unstructured repository for field notes and information related to the Maricopa field site. These 'raw' data can be imported into existing structured databases.

Manage Users

Edit Space

Extractors

Staging Area

Follow

External Links

Edit space to add links

Analysis Environment Instances

[David LeBauer's instance](#)

1 dataset uploaded

Refresh List

Environment Manager

Launch new instance with dataset

David LeBauer's instance

+ Launch


RStudio


Upload dataset to existing instance

Select an instance

+ Upload

Training

 This repository Search Pull requests Issues Marketplace Gist ToDo

terraref / tutorials 

Watch 26 Star 0 Fork 6


[Code](#) [Issues 9](#) [Pull requests 1](#) [Boards](#) [Reports](#) [Projects 0](#) [Wiki](#) [Insights](#)







Learn to use TERRA REF data and software [Edit](#)

[Add topics](#)

94 commits 4 branches 0 releases 8 contributors

Branch: master New pull request Create new file Upload files Find file [Clone or download](#)

 **dlebauer** committed on GitHub Merge pull request #20 from terraref/genomics Latest commit 50f1d4f2 6 days ago

 genomics	# open the key file	4 months ago
 plantcv	Added Clowder-PlantCV notebook	6 months ago
 sensors	minor changes to enable book to compile	3 months ago
 traits	merge	3 months ago
 workbench	Update ndslabs_workbench_intro.Rmd	5 months ago
 README.md	Update README.md	5 months ago

NDS Workbench



TERRA-REF Analysis Workbench

The TERRA-REF Analysis Workbench allows you to launch RStudio to analyze TERRA-REF data products. Data are currently available to our alpha user program.


[Learn more](#)


You are already signed in


 [Access your Configured Application Instances](#)


 [Browse the Application Catalog](#)


 [Add a New Application to the Catalog](#)


**File Manager**
Cloud Commander file manager and editor
Javascript Data transfer cloudcmd
View 1 Add


**Hyperspectral/NCO**
Hyperspectral workflow development environment
View 1 Add


**Jupyter/NetCDF4**
Jupyter SciPy Notebook with NetCDF support
Development environment Javascript Python
Analysis environment
View 1 Add


**Jupyter/PlantCV**
Jupyter SciPy Notebook with PlantCV support
Development environment Javascript Python
Analysis environment
Add

**Jupyter/TERRA**
Jupyter SciPy Notebook with TERRA dependencies
Development environment Javascript Python
Analysis environment
Add

**JupyterLab for pyCldower**
Extensible computational environment for Jupyter with pyCldower dependencies
Development environment Javascript Python
Analysis environment
View 1 Add

**NetCDF Visualization**
Xpra container for NetCDF Visualization
netcdfxpra
Add

**PostgresSQL Studio**
Web-based interface to PostgreSQL
pgstudio
View 1 Add

**RStudio Geospatial**
IDE for the R programming language.
Development environment Analysis environment R
geospatial
Add

<https://terraref.ndslabs.org>

Collaborators & Contact

David LeBauer dlebauer@illinois.edu

Website: terraref.org

GitHub: github.com/terraref

Twitter: [@terra_ref](https://twitter.com/terra_ref)

Donald Danforth Plant
Science Center



Todd Mockler
Project Lead



Nadia Shakoor
Associate Director



Noah Fahlgren
Phenotyping



Erica Fishel
Technology Transfer



Stuart Marshall
Scanalyzer Operator

University of Arizona,
Maricopa Agricultural Center



Pedro Andrade-Sanchez
Agronomy



Rick Ward
Field Phenotyping



Michael Ottman
Agronomy



Maria Newcomb
Field Coordinator

Saint Louis University



Abuduwasiti Wulamu
Remote Sensing

Texas A&M University



William Rooney
Breeding

George Washington University



Robert Pless
Image Analysis

USDA-ARS Arid Land
Agriculture Research Center



Jeff White
Agronomy

University of Illinois
at Urbana-Champaign



David LeBauer
Computing

Washington University
at St. Louis



Roman Garnett
Prediction Algorithms

Kansas State University



Geoff Morris
Gene-Trait Association



Jesse Poland
Phenotyping

LemnaTec GmbH



Solmaz Hajmohammadi
Image Analytics

National Center for
Supercomputing Applications



Max Burnette
Computing

HudsonAlpha Institute
for Biotechnology



Jeremy Schmutz
Sequencing

Clemson University



Stephen Kresovich
Breeding