



# **Phenotypic Data Collection, Management, and Availability**

# Overview

1. Introduction
2. Experiment Planning and Preparation
3. Data Collection
4. Data Management
5. Data Availability
6. Questions

# **Naser B. AlKhalifah**

## **Education:**

- M.S. Plant Breeding (Agronomy) – Iowa State University
- B.S. Genetics (Liberal Arts and Sciences) – Iowa State University

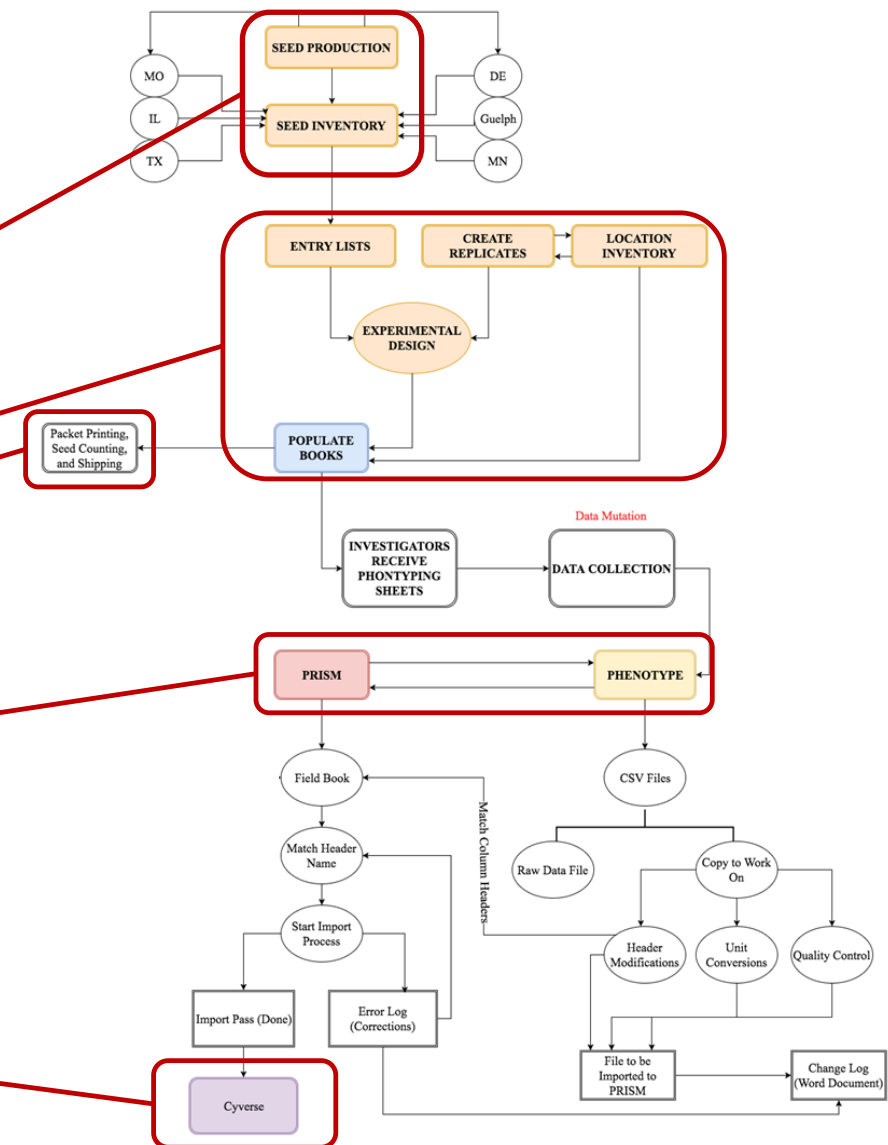
## **Professional Experience:**

- Assistant Scientist – Iowa State University (Current)
- Field Research Associate (II) – Monsanto Company (2014-2016)
- Lab Technician (II) – High-throughput Genotyping Lab: Monsanto Company (2009-2010)
- Seasonal Worker – Pioneer Hi-bred Intl. (2009)

# Introduction Cont'd

## Responsibilities:

- Planting locations and seed inventory
- Experiment planning & building
- Seed logistics
- Data curation
- Data release



# Overview

- ~~1. Introduction~~
2. Experiment Planning and Preparation
3. Data Collection
4. Data Management
5. Data Availability
6. Questions

# Experiment Planning and Preparation

1

Seed and location inventory received.

2

Fieldbooks and entry lists are created.

3

Seed packaging, shipping, and substitutions are made.

4

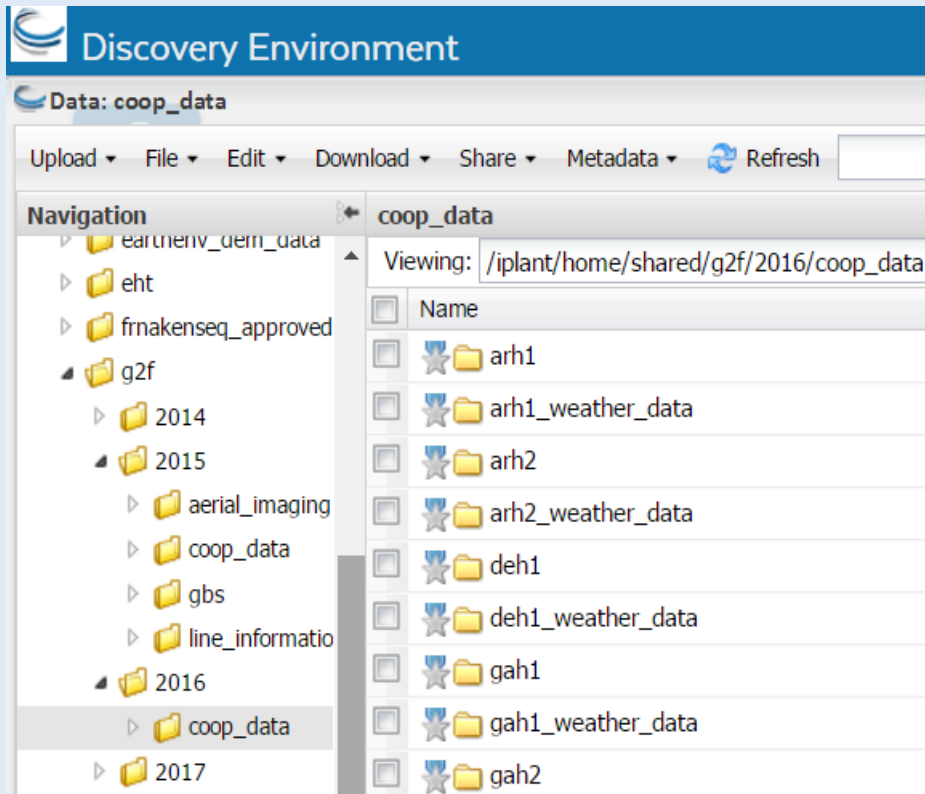
Updated fieldbooks are uploaded to both the Google Workbook and CyVerse

# Overview

- ~~1. Introduction~~
- ~~2. Experiment Planning and Preparation~~
3. Data Collection
4. Data Management
5. Data Availability
6. Questions

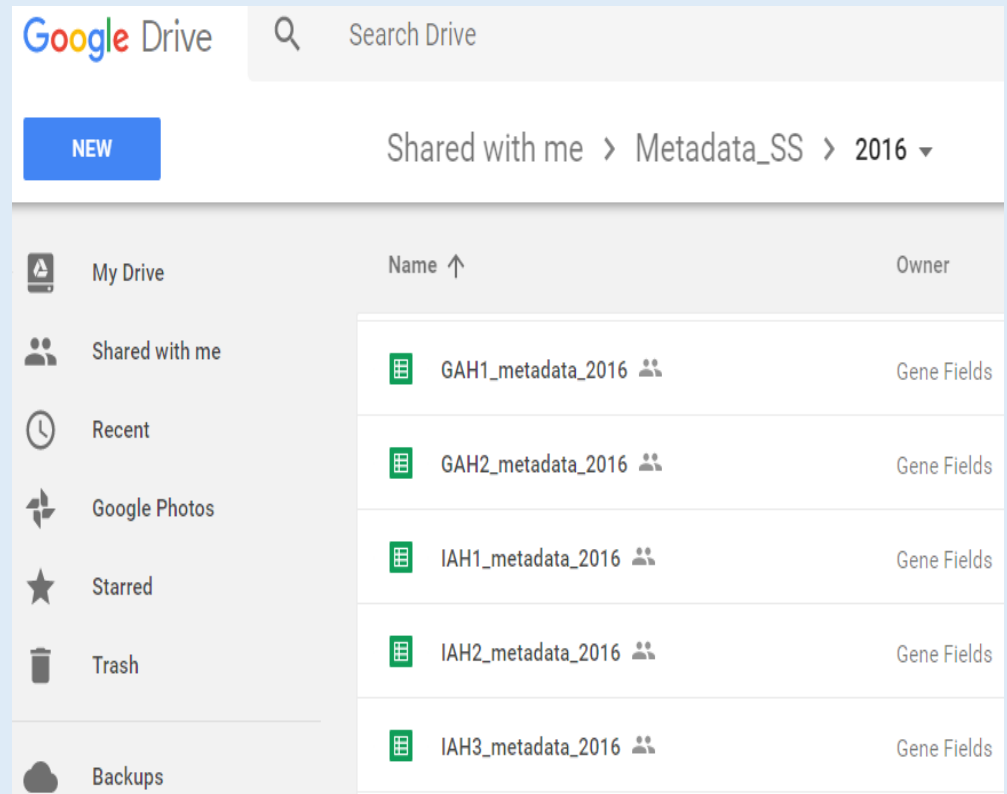
# Data Collection

## CyVerse



The screenshot shows the CyVerse Discovery Environment interface. The top bar is blue with the text "Discovery Environment". Below it, a breadcrumb trail reads "Data: coop\_data". A menu bar includes "Upload", "File", "Edit", "Download", "Share", "Metadata", and a "Refresh" button. The left sidebar, titled "Navigation", shows a tree structure of folders: "eartheniv\_uern\_data", "eht", "fnakenseq\_approved", "g2f", "2014", "2015", "2016", and "2017". The "2016" folder is selected, and its contents are listed in the main panel: "aerial\_imaging", "coop\_data", "gbs", "line\_informatio", and "coop\_data". The "coop\_data" folder is selected, and its contents are listed in the main panel: "arh1", "arh1\_weather\_data", "arh2", "arh2\_weather\_data", "deh1", "deh1\_weather\_data", "gah1", "gah1\_weather\_data", and "gah2".

## Google Sheets



The screenshot shows the Google Drive interface. The top bar is white with the text "Google Drive" and a search bar labeled "Search Drive". Below the bar, a blue "NEW" button is on the left, and a breadcrumb trail reads "Shared with me > Metadata\_SS > 2016". The left sidebar shows a list of folders: "My Drive", "Shared with me", "Recent", "Google Photos", "Starred", "Trash", and "Backups". The main panel displays a list of spreadsheets for the year 2016, sorted by name. The list includes:

| Name               | Owner       |
|--------------------|-------------|
| GAH1_metadata_2016 | Gene Fields |
| GAH2_metadata_2016 | Gene Fields |
| IAH1_metadata_2016 | Gene Fields |
| IAH2_metadata_2016 | Gene Fields |
| IAH3_metadata_2016 | Gene Fields |



# Overview

- ~~1. Introduction~~
- ~~2. Experiment Planning and Preparation~~
- ~~3. Data Collection~~
- 4. Data Management
- 5. Data Availability
- 6. Questions

# Data Management

Visually inspect csv files for data type inconsistencies

|    | K                      | L                    | M                 | N               | O                         |
|----|------------------------|----------------------|-------------------|-----------------|---------------------------|
| 1  | Pollen Date [MM/DD/YY] | Silk Date [MM/DD/YY] | Plant Height [cm] | Ear Height [cm] | Stand Count [# of plants] |
| 2  | July-10-16             | 7/9/2016             | 117               | 270             | 57-some snapped           |
| 3  | July-10-16             | 7/9/2016             | 269               | 142             | 61                        |
| 4  | 7/9/2016               | 7/10/2016            | 254               | 116             | Missing                   |
| 5  | 7/11/2016              | 7/11/2016            | 227               | 102             | 57                        |
| 6  | 7/12/2016              | 7/11/2016            | 234               | 123             | 46                        |
| 7  |                        | 7/8/2016             | 260               | 113             | 59                        |
| 8  | 7/6/2016               | 7/6/2016             | 250               | 111             | 53                        |
| 9  | 7/8/2016               | 7/9/2016             | 274               | 149             | 53                        |
| 10 | 7/10/2016              | 7/9/2016             | 278               | 150             | 59                        |
| 11 | 7/6/2016               | 7/6/2016             | 271               | 132             | 56                        |
| 12 | 7/7/2016               | 7/5/2016             | 275               | 149             | 60                        |
| 13 | 7/7/2016               | 7/7/2016             | 275               | 150             | 55                        |
| 14 | 7/8/2016               | 7/8/2016             | 254               | 119             | 46                        |
| 15 | 7/9/2016               | 7/9/2016             | 240               | 111             | 57                        |
| 16 | 7/6/2016               | 7/6/2016             | 239               | 92              | 57                        |
| 17 | 7/7/2016               | 7/8/2016             | 272               | 127             | 58                        |
| 18 | 7/6/2016               | 7/6/2016             | 237               | 108             | 56                        |
| 19 | 7/10/2016              | 7/9/2016             | 267               | 132             | 56                        |


# Data Management

## Import data files to PRISM software

PRISM - Naser Alkhalifah

File Edit Related Books Procedures Format Data Window Help

Book Explorer Yield Trial Master Catalog x



|   | Date Planted | Date Harvested | Pollen Date | Silk Date | STAN D | Polle n DA | Silk DAP | PLTHT | EARH T | RTL DG | SKL DG | MOIST | TWT  | WT    |
|---|--------------|----------------|-------------|-----------|--------|------------|----------|-------|--------|--------|--------|-------|------|-------|
|   | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/13/2016 | 61     | 67         | 67       | 170   | 108    |        |        | 16.4  | 61.2 | 15.99 |
|   | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/15/2016 | 95     | 67         | 69       | 178   | 123    |        | 1      | 14.6  | 62.9 | 20.33 |
|   | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/15/2016 | 79     | 67         | 69       | 173   | 104    |        | 2      | 14.5  | 62.8 | 20.26 |
|   | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/15/2016 | 77     | 67         | 69       | 175   | 121    |        | 6      | 14.7  | 62.6 | 18.45 |
| > | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/15/2016 | 78     | 67         | 69       | 169   | 121    |        | 3      | 14.9  | 62.3 | 20.57 |
|   | 4/7/2016     | 8/30/2016      | 6/13/2016   | 6/15/2016 | 80     | 67         | 69       | 183   | 130    |        |        | 15.7  | 62.0 | 21.15 |
|   |              | 8/30/2016      | 6/15/2016   | 6/17/2016 | 94     |            |          | 199   | 94     |        |        | 17.2  | 60.9 | 31.05 |
|   |              | 8/30/2016      | 6/15/2016   | 6/16/2016 | 95     |            |          | 188   | 72     |        |        | 17.1  | 61.4 | 30.55 |
|   |              | 8/30/2016      | 6/13/2016   | 6/13/2016 | 90     |            |          | 175   | 88     |        |        | 14.8  | 62.8 | 23.14 |
|   |              | 8/30/2016      | 6/15/2016   | 6/16/2016 | 82     |            |          | 180   | 89     |        |        | 15.3  | 62.4 | 27.14 |
|   |              | 8/30/2016      | 6/15/2016   | 6/16/2016 | 88     |            |          | 179   | 85     |        | 2      | 15.9  | 62.3 | 19.46 |

# Data Management

## Conditions for declaring observations out of range.

| Trait             | Condition for Setting to Missing | Columns Set to Missing |
|-------------------|----------------------------------|------------------------|
| Pollen DAP [days] | Less than 20 or more than 100    | Pollen DAP             |
| Silk DAP [days]   | Greater than 100                 | Silk DAP               |
| Ear Height [cm]   | Less than 20                     | Ear Height.            |

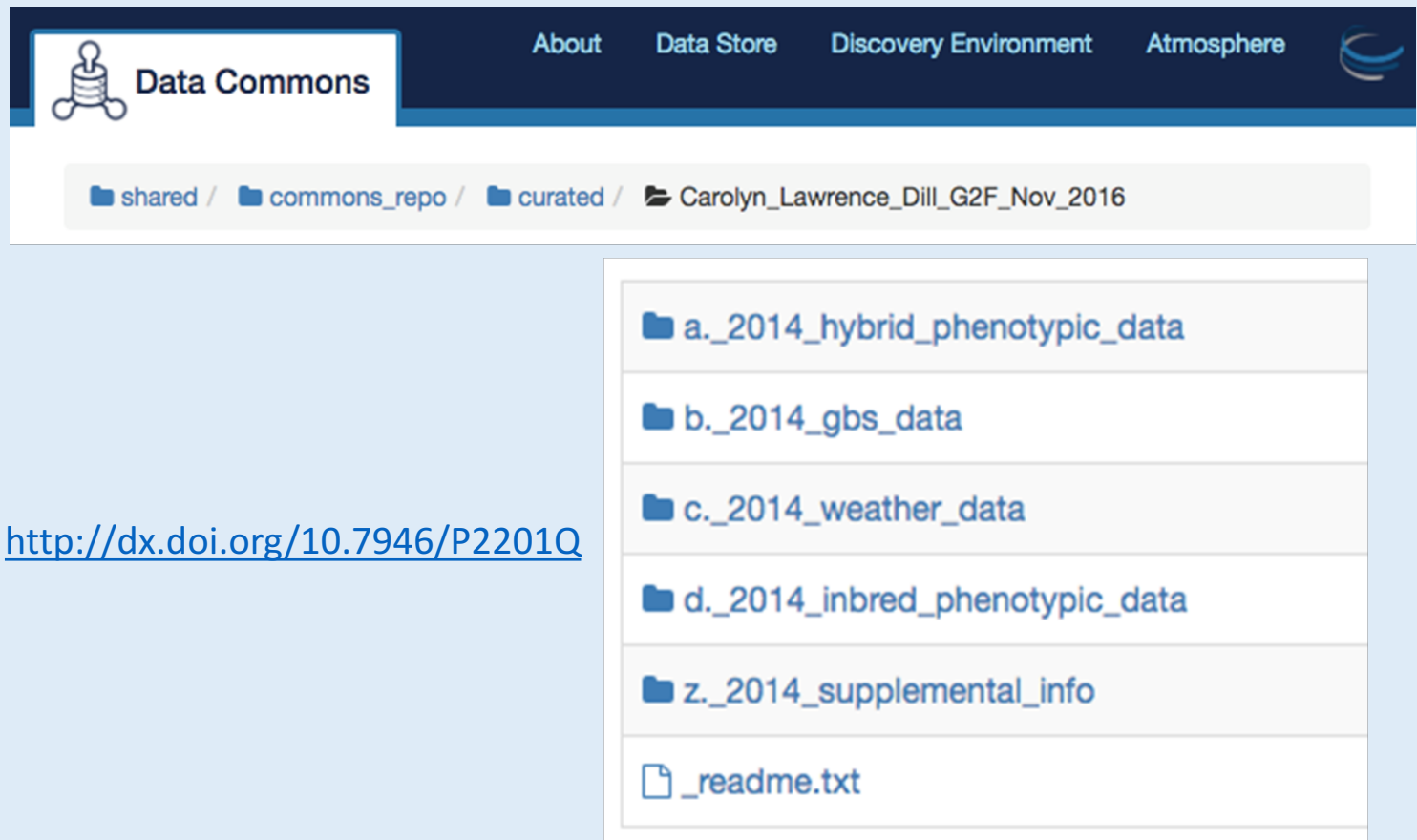
# Overview

- ~~1. Introduction~~
- ~~2. Experiment Planning and Preparation~~
- ~~3. Data Collection~~
- ~~4. Data Management~~
- 5. Data Availability
- 6. Questions

# Data Availability

| Field Season | Release to Collaborators (ARK) | Release to Public (DOI)   |
|--------------|--------------------------------|---|
| 2014         | -- NA --                       | <a href="http://dx.doi.org/10.7946/P2201Q">http://dx.doi.org/10.7946/P2201Q</a> |
| 2015         | February 2017                  | <a href="http://dx.doi.org/10.7946/P24S31">http://dx.doi.org/10.7946/P24S31</a> |
| 2016         | February 2017                  | March 2018  |
| N            | February of N+1                | March of N+2  |

# Data Availability



The screenshot displays the Data Commons website. The top navigation bar includes links for 'About', 'Data Store', 'Discovery Environment', and 'Atmosphere'. Below the navigation bar, a breadcrumb trail shows the path: 'shared / commons\_repo / curated / Carolyn\_Lawrence\_Dill\_G2F\_Nov\_2016'. The main content area lists several data folders and one text file:

- folder a.\_2014\_hybrid\_phenotypic\_data
- folder b.\_2014\_gbs\_data
- folder c.\_2014\_weather\_data
- folder d.\_2014\_inbred\_phenotypic\_data
- folder z.\_2014\_supplemental\_info
- file \_readme.txt

DOI for 2014: <http://dx.doi.org/10.7946/P2201Q>

# G X E Cooperators

*Principal Investigators who grew GxE trials in 2014-2016*

- |                              |                              |                              |
|------------------------------|------------------------------|------------------------------|
| • Martin Bohn (UIUC)         | ✧ Joe Knoll (ARS - UGA)      | ✧ Oscar Rodriguez (UNL)      |
| • Ed Buckler (ARS - Cornell) | ✧ Judith Kolkman (Cornell)   | ✧ Cinta Romay (Cornell)      |
| • Ignacio Ciampitti (KSU)    | ✧ Greg Kruger (UNL)          | ✧ James Schnable (UNL)       |
| • Jode Edwards (ARS - ISU)   | ✧ Nick Lauter (ARS - ISU)    | ✧ Pat Schnable (ISU)         |
| • Sherry Flint-Garcia (ARS)  | ✧ Liz Lee (Guelph)           | ✧ Brian Scully (ARS)         |
| • Christopher Graham (SDSU)  | ✧ Natalia de Leon (UW)       | ✧ Rajandeep Sekhon (Clemson) |
| • Mike Gore (Cornell)        | ✧ Sanzhen Liu (KSU)          | ✧ Margaret Smith (Cornell)   |
| • Candy Hirsch (UMN)         | ✧ Argelia Lorence (AR-State) | ✧ Nathan Springer (UMN)      |
| • Jim Holland (ARS - NCSU)   | ✧ Aaron Lorenz (UMN)         | ✧ Kurt Thelen (MSU)          |
| • Elizabeth Hood (AR-State)  | ✧ Jonathan Lynch (PSU)       | ✧ Peter Thomison (OSU)       |
| • David Hooker (Guelph)      | ✧ Steve Moose (UIUC)         | ✧ Mitch Tuinstra (Purdue)    |
| • Fiona Goggin (Univ AR)     | ✧ Seth Murray (TAMU)         | ✧ Jason Wallace (UGA)        |
| • Shawn Kaeppler (UW)        | ✧ Rebecca Nelson (Cornell)   | ✧ Randy Wisser (UDel)        |
|                              | ✧ Torbert Rocheford (Purdue) | ✧ Wenwei Xu (TAMU)           |





# Genomes To Fields Collaborators

## † G2F Executive Committee members

- ✧ Naser Alkhalifah (ISU)
- ✧ Martin Bohn (UIUC)\*
- ✧ Ed Buckler (ARS)†
- ✧ Darwin Campbell (ISU)\*
- ✧ Ignacio Ciampitti (KSU)
- ✧ James Clohessy (Cornell)
- ✧ Liang Dong (ISU)
- ✧ Jode Edwards (ARS)\*
- ✧ David Ertl (IA Corn)\*†
- ✧ Sherry Flint-Garcia (ARS)\*
- ✧ Joseph Gage (UW)
- ✧ Jack Gardiner (ISU)\*
- ✧ Fiona Goggin (Univ AR)
- ✧ Byron Good (Guelph)
- ✧ Mike Gore (Cornell)
- ✧ Christopher Graham (SDSU)
- ✧ Patricio Grassini (UNL)
- ✧ Jerry Hatfield (ARS)
- ✧ Candy Hirsch (UMN)
- ✧ Jim Holland (ARS)
- ✧ Elizabeth Hood (AR-State)
- ✧ David Hooker (Guelph)

## \*GxE Coordinating Groups

- ✧ Diego Jarquin (UNL)\*
- ✧ Shawn Kaeppler (UW)†
- ✧ Joe Knoll (ARS)
- ✧ Judith Kolkman (Cornell)
- ✧ Greg Kruger (UNL)
- ✧ Nick Lauter (ARS)
- ✧ Carolyn Lawrence-Dill (ISU)\*
- ✧ Liz Lee (Guelph)\*
- ✧ Zhizhai Liu (TAMU)
- ✧ Natalia de Leon (UW)\*†~
- ✧ Alex Lipka (UIUC)
- ✧ Argelia Lorence (AR-State)
- ✧ Aaron Lorenz (UMN)\*
- ✧ Jonathan Lynch (PSU)†
- ✧ Nathan Miller (UW)
- ✧ Steve Moose (UIUC)
- ✧ Seth Murray (TAMU)\*
- ✧ Rebecca Nelson (Cornell)
- ✧ Torbert Rocheford (Purdue)
- ✧ Oscar Rodriguez (UNL)
- ✧ Cinta Romay (Cornell)\*
- ✧ Emily Rothfusz (UW)\*

## ~G2F co-lead

- ✧ James Schnable (UNL)
- ✧ Pat Schnable (ISU)†~
- ✧ Brian Scully (ARS)
- ✧ Rajandeep Sekhon (Clemson)
- ✧ Kevin Silverstein (UMN)
- ✧ Margaret Smith (Cornell)
- ✧ Edgar Spalding (UW)
- ✧ Nathan Springer (UMN)†
- ✧ Srikant Srinivasan (ISU)\*
- ✧ Yiwei Sun (ISU)\*
- ✧ Kurt Thelen (MSU)
- ✧ Peter Thomison (OSU)
- ✧ Kelly Thorp (ARS)
- ✧ Mitch Tuinstra (Purdue)
- ✧ Jason Wallace (UGA)
- ✧ Renee Walton (ISU)
- ✧ Rod Williamson (IA Corn)
- ✧ Randy Wissner (UDel)\*
- ✧ Wenwei Xu (TAMU)
- ✧ Cheng-Ting Yeh (ISU)
- ✧ Jianming Yu (ISU)



## Genomes To Fields Sponsors



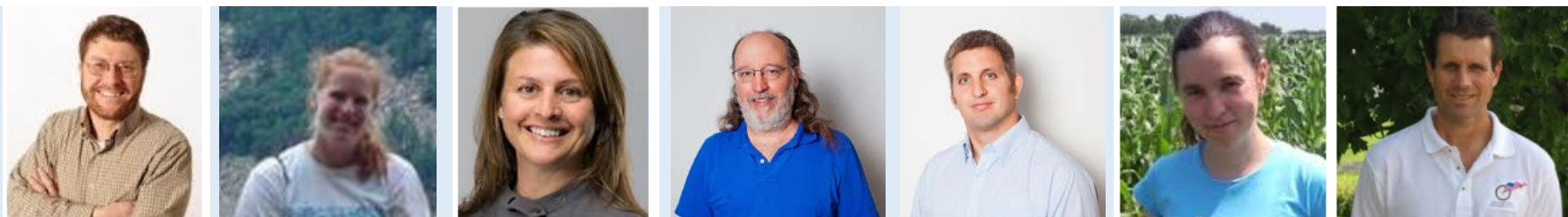
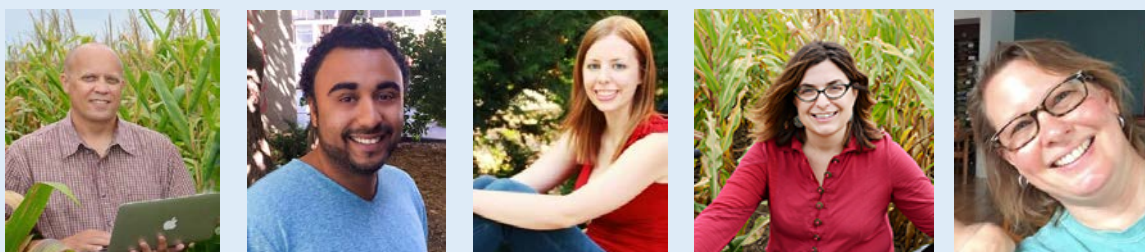
United States Department of Agriculture  
National Institute of Food and Agriculture





## Phenotypic Data Collection, Management, and Availability

**IOWA STATE UNIVERSITY**  
Plant Sciences Institute



**Thank you for your time!**

Questions?