Proprietary Data in Data-Driven Research

Internet2 Technology Exchange 23

Jesse Erdmann

University of Minnesota
Driven to Discover℠

GEMS Informatics

GENOMICS | ENVIRONMENT | MANAGEMENT | SOCIO-ECONOMICS

CFANS
COLLEGE OF FOOD, AGRICULTURAL AND NATURAL RESOURCE SCIENCES

Minnesota Supercomputing Institute
University of Minnesota
Driven to Discover®

U-SPATIAL

University of Minnesota
Driven to Discover℠

Public-Private Partnerships

7 Public-Private Research Consortia

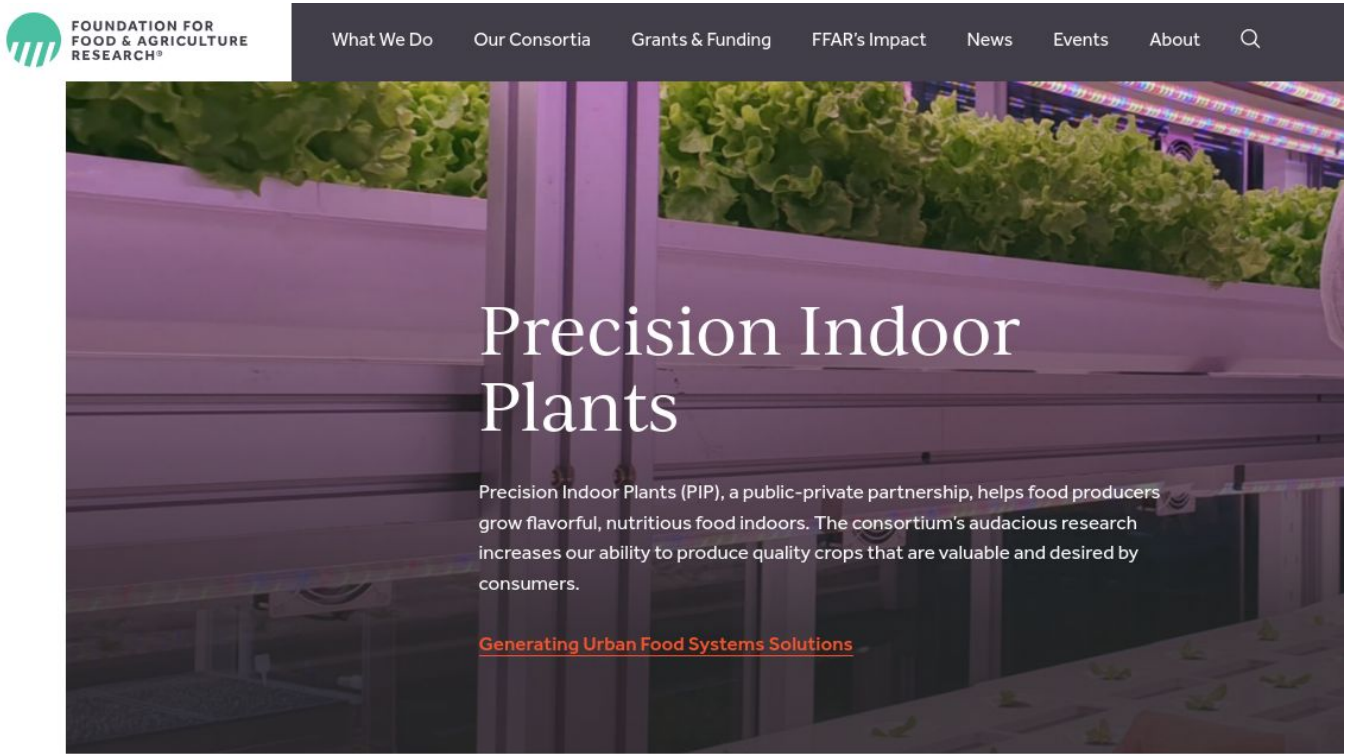
- Drought tolerant corn
- Heat tolerant field tomatoes
- Pea protein quality
- Indoor lettuce quality, shelf life
- Indoor tomato flavor
- Bean breeding bottlenecks
- Potato, citrus insect resistance

The challenges in public/private partnerships

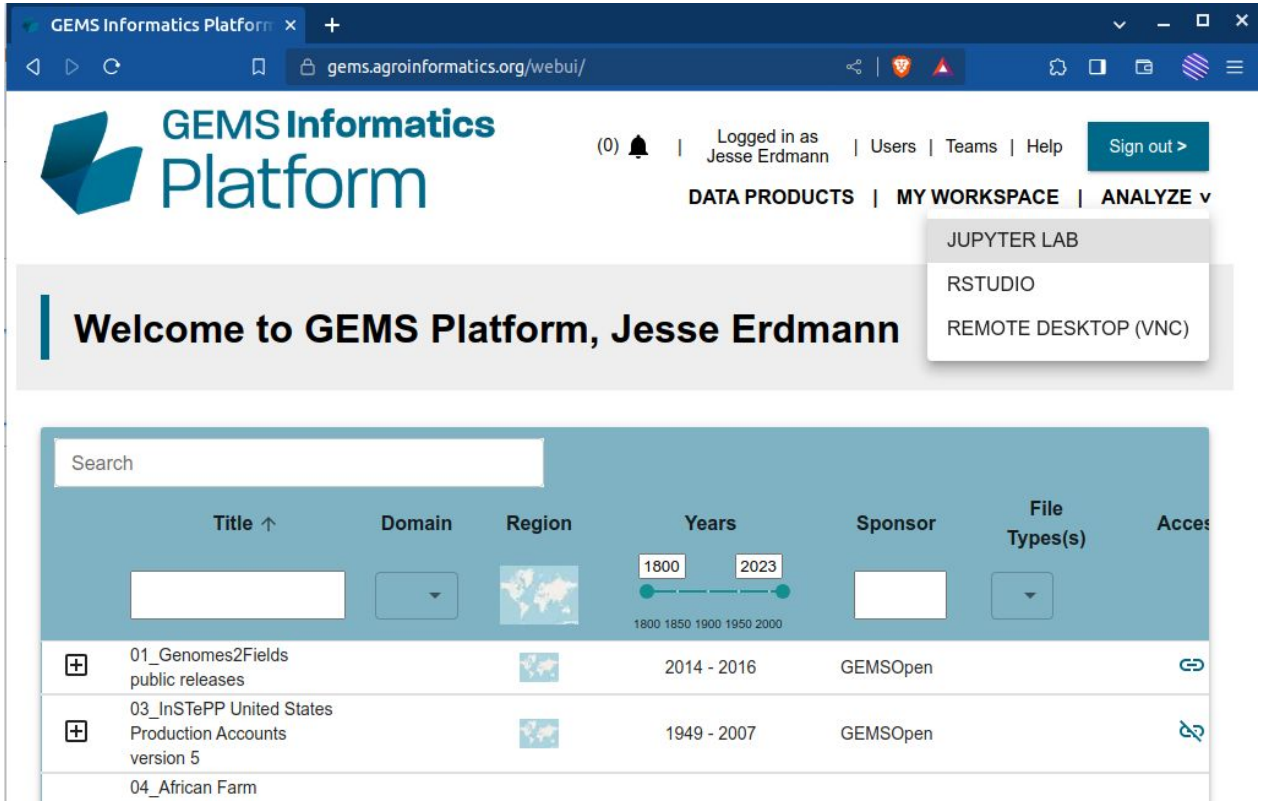
- In general…
  - More data, more variety
  - Skills gap
- In academia…
  - Transient student workforce
  - Incentives de-prioritize skill acquisition for many academics
  - Money is scarce and dwindling
- In agriculture…
  - Data often includes personally identifiable information
  - Companies tend to be very focused on data protection
  - Companies dramatically outspend academia on research

An example project…

- oAuth2 and data transfer via Globus
- Metadata capture using ontological terms
- Private users share with central, public org
- Public org cleans, aggregates, and publishes back to community



Addressing the skills gap with tools…



- Programming environments prebuilt with common tools
- Containerized environments to aid reproducibility

Learning

THE NEXT GENERATION OF AGRI-FOOD DATA SCIENTISTS

GEMS Learning provides sensing and data science education tailored to food, agriculture, and natural resource applications. Customized short courses serve different needs with particular emphasis on recruitment and funding of traditionally disadvantaged groups.

GEMS LEARNING FEATURES

- State-of-the-art data science
- Customized modules for the agri-food sector
- Class projects rooted in agri-food use cases and data
- Classes at all levels, even aspiring data scientists

FIND OUT MORE ONLINE

GEMS Learning

INDIVIDUALIZED AGRI-FOOD DATA SCIENCE MODULES AND PATHWAYS

× Computing Basics for the Agri-food Sector

Are you a field or bench scientist and always wanted to feel more comfortable with your computing skills? These courses are designed for those who have never used the command line, but realize that the responsibilities they have or will soon take on require them to automate tasks. It will teach basic UNIX command-line skills, enable participants to work remotely on more powerful machines, create and run scripts to automate complex workflows, and synchronize your scripts with the larger community with Github.

- Introducing the GEMS platform + Jupyter Lab
- Demystifying the UNIX command line
- Working Remotely and Scheduling Jobs on MSI's systems
- Synching your work with the community

+ Introduction to Data Analysis with R

+ Accounting for Location in Agriculture in R

+ Accounting for Location in Agriculture in Python

+ Spatio-Temporal Accounting of Biotic Threats

+ Digital Agriculture

+ Topics Courses

Privacy...

- User visibility in the platform
- Tools
  - Geospatial fuzzing
  - Geospatial or value based aggregation
  - Identifiable field substitution
- and...

Legislative/Legal Privacy

BE IT ENACTED BY THE LEGISLATURE OF THE STATE OF MINNESOTA:

Section 1. Minnesota Statutes 2016, section 13.643, subdivision 7, is amended to read:

Subd. 7. **Research, monitoring, or assessment data.** (a) Except as provided in paragraph (b), the following data created, collected, ~~and~~ or maintained by the Department of Agriculture or the University of Minnesota during research, monitoring, or the assessment of farm practices and related to natural resources, the environment, agricultural facilities, or agricultural practices are classified as private or nonpublic:

(1) names, addresses, telephone numbers, and e-mail addresses of study participants or cooperators; ~~and~~

(2) location of research, study site, and global positioning system data; an

(3) data created, collected, or maintained by the University of Minnesota for inclusion on an agricultural data analysis platform maintained and hosted by the University of Minnesota that identify or could identify an individual or business.

Law came into effect

August 1, 2018

Future technology integrations…

The Dataverse® Project

iRODS®

Thanks

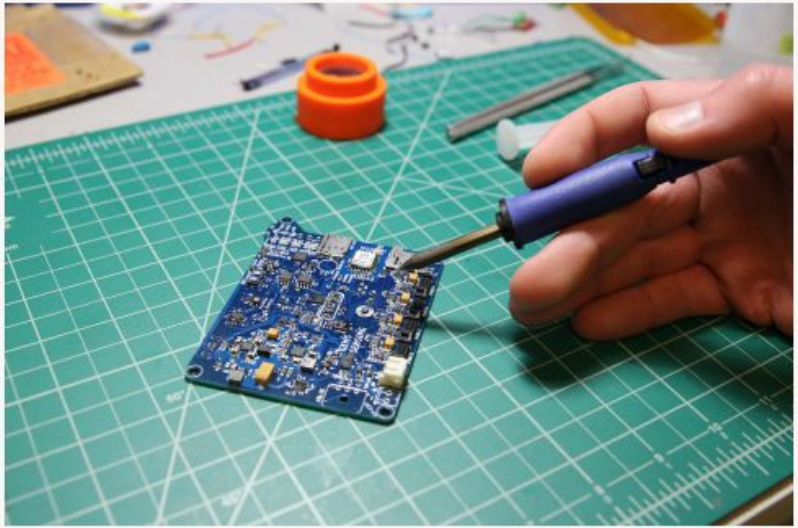
https://gems.umn.edu

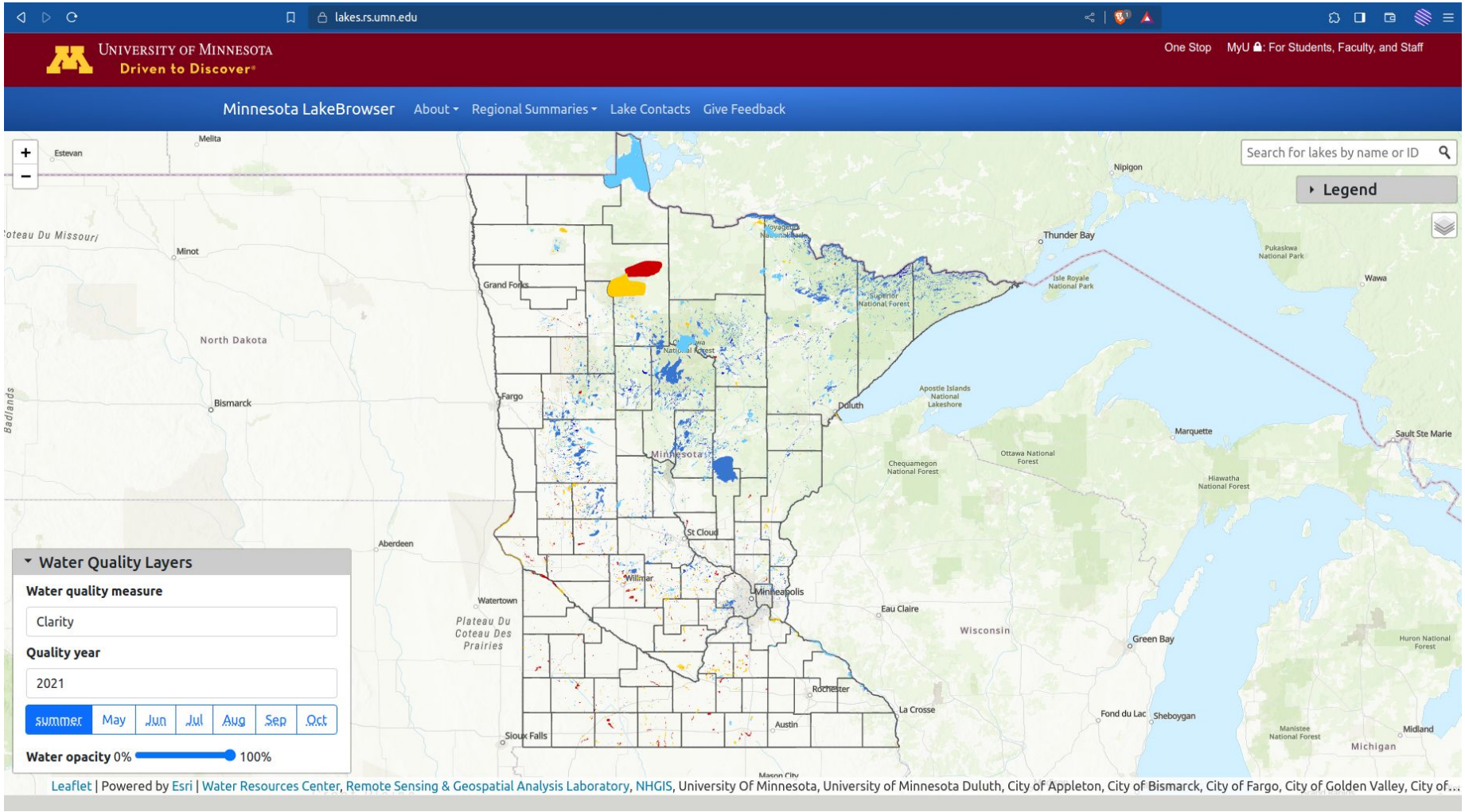
GEMS Platform | GEMS Exchange | GEMS Sensing | GEMS Learning | GEMS Solutions

GEMS **SENSING**

Making it as easy and cheap as possible for scientists to collect real-time, standardized, and quality assured field data.

View More about GEMS Sensing ›




lakes.rs.umn.edu
University of Minnesota
Driven to Discover
One Stop
MyU: For Students, Faculty, and Staff
Minnesota LakeBrowser
About
Regional Summaries
Lake Contacts
Give Feedback
Search for lakes by name or ID
Legend
Water Quality Layers
Water quality measure
Clarity
Quality year
2021
summer
May
Jun
Jul
Aug
Sep
Oct
Water opacity 0%
100%
Leaflet | Powered by Esri | Water Resources Center, Remote Sensing & Geospatial Analysis Laboratory, NHGIS, University Of Minnesota, University of Minnesota Duluth, City of Appleton, City of Bismarck, City of Fargo, City of Golden Valley, City of...