Building an Agriculture Data Infrastructure for Arizona

Barney Maccabe, Professor

HARVEST Workshop
September 10, 2025
San Diego, CA

THE UNIVERSITY OF ARIZONA
College of Information Science

We respectfully acknowledge the University of Arizona is on the land and territories of Indigenous peoples. Today, Arizona is home to 22 federally recognized tribes, with Tucson being home to the O'odham and the Yaqui. Committed to diversity and inclusion, the University strives to build sustainable relationships with sovereign Native Nations and Indigenous communities through education offerings, partnerships, and community service.

THE UNIVERSITY OF ARIZONA
College of Information Science

Stargate

- Announced January 21, 2025: LLC OpenAI, SoftBank, Oracle, and MGX
- $500B investment in AI infrastructure in the US by 2029
  - $100B in 2025
  - "**Manhattan Project**" for AI (Industry exclusive)
- Federal role: remove regulatory barriers
- In July, DOE announced four AI Data Center Sites
  - ORNL, Idaho, Paducah, and Savannah River
  - Executive order removes clean energy rules and streamlines reviews to accelerate deployment

- Labs and universities – integration of instruments
  - NSF PCL & DOE IRI

West Texas (Abilene) Data Center
- 875 acres (bigger than Central Park)
- 1.2 GW + 4.5 GW (Oracle)
- $7.1 billion loan led by JPMorgan

Arizona Economy: 5 C's

- Copper
- Cattle
- Cotton (Pima)
- Citrus
- Climate

Arizona Economy: 5 C's

Caterpillar, a manufacturer of construction and mining equipment, has built a prototype for an all electric truck. The 240-ton large mining truck was demonstrated in November 2022. The truck is on display at Caterpillar's Tucson Proving Ground in Green Valley.

Mamta Popat, Arizona Daily Star

- Copper
- Cattle → Cows (dairy)
- Cotton (Pima) → Computing?
- Citrus → Crops (leafy greens)
- Climate "*We have the climate of tomorrow, today.*"
  –Duke Pauli

THE UNIVERSITY OF ARIZONA
College of Information Science

Agriculture in AZ

- $23.3B Economic Impact
- Supports 138,000+ jobs
- 20,000 farms/ranches (26M acres)
- Exports to over 70 countries

3 of Arizona's "Five C's"

Source: nass.usda.gov/az/

"Arizona is known as one of the most productive and efficient agricultural regions in the world and one of the most diverse agricultural production states in the nation."

AZ Farm Bureau

Slide courtesy of Connor Osgood

THE UNIVERSITY OF ARIZONA
College of Information Science

Ag in Arizona (Yuma) is Technology Intensive

Yuma
Cyber Experiment Station
YCEDA (Yuma Center for Excellence in Desert Agriculture)
Examples

THE UNIVERSITY OF ARIZONA
College of
Information Science

Agriculture in Yuma

- High value winter specialty crop industry ($2B)
- Supplies ~80% of the winter leafy green vegetables for USA and Canada (Nov – Mar)
- Top 0.1% of U.S. Counties in vegetable sales, Top 0.5% for all agricultural products
- One of the top worldwide producing areas of medjool dates (7,000+ acres)
- Year-round growing season, 150+ crops
  - Rotation crops: Wheat, Melons, Cotton, Hay
  - Also: Citrus and Seed Crops

- 230,000 acres of fertile soil
- Available workforce
- Senior water rights
- Great community support
- Progressive farmers exploring/utilizing technology

Slide courtesy of Connor Osgood

THE UNIVERSITY OF ARIZONA
Yuma Center of Excellence for Desert Agriculture

THE UNIVERSITY OF ARIZONA
Arizona Experiment Station

Cyber Experiment Station

- Team of data scientists, programmers/developers, cloud-solution experts, videographers, and telepresence engineers.
- Provides the digital-communications vehicle to deliver our “story”.
  ½ Comms + ½ Cyber Technologies
- Unique within the National Agricultural Experiment Station network

THE UNIVERSITY OF ARIZONA
Yuma Center of Excellence for Desert Agriculture

YCEDA

- YCEDA is an innovative public-private partnership that connects top scientists to the desert agricultural industry.
- Together we put science to work developing solutions to the challenges of arid-land crop production.
- Our work focuses on high-priority issues identified by industry stakeholders, including but not limited to increasing production efficiencies through disease and water management, crop yield maximization, food safety, and technology utilization.

The Desert Difference: FarmTech Connect

- Day 1 November 13
  Practical Demonstrations at the Yuma Agricultural Center
  - Field Day: 7:00 AM – 12:00 PM
  - Lunch provided: 12:00 – 1:00 PM
- Day 2 November 14
  Conference Panels, Presentations, and Discussions
  - 8:00 AM — 12:00 PM Panels, speakers and pitches
- https://desertagsolutions.org/events/desert-difference-farmtech-connect

THE UNIVERSITY OF ARIZONA
College of Information Science

Ex. 1: Robotic crop thinning & weeding

Vision Robotics | RDO website

University of Arizona | FarmProgress

- Lettuce seeds are planted with 2” spacing, crop is thinned to 12” spacing
  - Helps to guarantee ‘a plant in every spot’
- Stand count and plant uniformity for is critical for profitability in lettuce
- Thinning is done manually by contract labor crews (typically H2A)
- While crews work quickly, there are issues with consistency
- Robotic lettuce thinning & cultivation
  - Image recognition to target plants for removal
  - Precision spot spray of herbicide or lasers to kill extra plants
  - One company claims 7 acres/hour

Slide courtesy of Connor Osgood

Ex. 2: Irrigation and fertilizer management

Water content vs. water potential (METER Group)

Irrometer

TEROS 21 | METER Group

RS130 Telemetry Node +

- Delivering water to the crop based on what is needed, not on time schedule
- Advanced sensors
- Remote/proximal crop sensing
- Crop water stress index (CWSI)
- Nutrient-specific sensors (some day…)
- Variable-rate fertilizer based on local soil maps
- Validation studies needed to ensure accuracy of models

Slide courtesy of Connor Osgood

THE UNIVERSITY OF ARIZONA
College of Information Science

Incremental Experimentation

Controlled Environments

Biosphere 2

The Gantry

THE UNIVERSITY OF ARIZONA
College of Information Science

Controlled Environment Agriculture Center

Controlled Environments

- University of Arizona Controlled Environment Agriculture Center (CEAC) supports research related to plant growth in alternate environments
- Upper right image shows the Prototype Lunar Greenhouse
- Lower right shows agrovoltaics where part of the incident solar radiation not required for crop growth is harvested for generating power

Biosphere 2: An Analog Twin

- The University of Arizona's Biosphere 2 is the world's largest controlled environment dedicated to understanding the implication, mitigation and adaptation solutions for resilience of our planet (Biosphere 1) due to the global climate crisis.
- Biosphere 2 is a meso-scale Earth science facility encompassing 3.14 acres and houses five synthetic ecosystems encased in a glass and metal spaceframe. Operational for 30 years and counting, these meso-scale ecosystems include the world's largest controlled systems of tropical rain forest, desert, savanna, mangrove, and ocean.

… and a Digital Twin

Using Nvidia's Omniverse

Run virtual experiments before physical experiments

THE UNIVERSITY OF ARIZONA
College of Information Science

Big Data Problem: Phenotyping 30K Individual Plants

- University of Arizona is home to the world's largest, outdoor phenotyping platform – the Field Scanner.
- Provides unique opportunity to study how plants grow and respond to extreme environmental conditions, like those predicted in the future climate scenarios.
- Collecting and providing high resolution data that has never been available to the research community.

THE UNIVERSITY OF ARIZONA
College of Information Science

Sensors on the scanalyzer

RGB
~550GB

Thermal
~5.5GB

Fluorescence
~80GB

3D
~300GB

Hyperspectral
~3.2TB

Creation of sorghum digital twins

Data acquisition

Point clouds

Digital twin

“Plot” of digital twins for study

Field Data

Arizona Broadband Strategy

Youm County last-mile for agriculture

THE UNIVERSITY OF ARIZONA
College of
Information Science

AZ Broadband Strategic Plan: Build the Middle Mile

- 80% of unserved and underserved households can be served by the middle-mile network on interstate and select state highways
- The middle-mile network creates the open access network that last mile initiatives can tap into to cover the remaining 21% of underserved

THE UNIVERSITY OF ARIZONA
College of Information Science

Slide courtesy of Connor Osgood

Connecting fields in Yuma County

Image courtesy of Connor Osgood

- NSF CC* grant; connectivity:
  - Experiment station
  - Local education facilities
- Longer term Yuma Board of Supervisors
- 27 CBRS towers
  - 10GB up and download
  - ~180,000 acres, ~450 farms
- Three tiers of service:
  - Wi-Fi,
  - Broadband cellular (CBRS), and
  - LoRaWAN for battery and solar devices

THE UNIVERSITY OF ARIZONA
College of Information Science

Yuma updates

- Upgrade 4G LTE radios at the Yuma Ag Center to 5G at the end of August improving the speed of in-the-field connected devices.
- We have developed a LoRaWAN IoT dashboard for all Arizona Experiment Stations (AES).
- Deploying FLOTO network monitoring nodes in and around our fields at our Yuma Ag Center.
  - Data collected from these devices will be included in a nationwide broadband diagnostic testing platform – more at https://floto.cs.uchicago.edu/about/
- The Sun Corridor Network has installed two PerfSonar nodes on the 10G/100G backhaul fiber network; one at the Yuma Ag Center and another at the 100G POP at Yuma Union High School. We'll be able to track network performance, collect statistics, identify peak load times, possibly identify application use, and measure network throughput between the Yuma Ag Center and CyVerse.
- A November "Save-the-Date" will be going out soon announcing the ribbon cutting event, unveiling the first Yuma Broadband Authority-managed cellular network tower, prioritized for agriculture.

Integration: Connecting Data Sources

Making the whole greater than the sum of the parts

Cautions:

- More data isn't always good
- Data always has bias
- Gaps in data need to be filled

THE UNIVERSITY OF ARIZONA
College of Information Science

Integration

Making the whole greater than the sum of the parts

- Experiments spanning multiple scales
- Digital twins for “in silico” experiments
- Building better models

Fields in Yuma

The Gantry

Biosphere 2

Controlled Environment

Cyber Experiment Station

CYVERSE®

THE UNIVERSITY OF ARIZONA
College of Information Science

Caution 1: More data isn't always good

- The Deluge of Spurious Correlations in Big Data
  - Calude, C.S., Longo, G., Foundations of Science, March 2016 (https://doi.org/10.1007/s10699-016-9489-4)
  - More data leads to more spurious correlations

Must be intentional in the collection and aggregation of data

THE UNIVERSITY OF ARIZONA
College of Information Science

Caution 2: Data has value and bias

- Collecting data is expensive: Data has value
  - Data Owners must be compensated for use

Need better tools for attribution

- Reuse is dangerous: All data has bias
  - Metadata will never capture the full context for data, Data Owners know the full context

How much is enough?

Data owners must be engaged in the use of their data

Related to Data Privacy and Data Sovereignty

Caution 3: Gaps must be filled

- Aggregated data sets will have holes in the sample space
- Science prefers exploration of "interesting" parts of the sample space
  - This is especially true for incremental science
- Models need to be trained on the whole space

Funding needs to support exploration of the entire sample space:
Autonomous instruments may help

A closing example: MDRepo

A Decentralized Repository for MD Simulations

THE UNIVERSITY OF ARIZONA
College of
Information Science

MDRepo: A Community Resource for MD Simulations

- Molecular Dynamic simulations (e.g., LAMMPS) consume significant computing resource
- Build a community repository to share results and avoid redundant computation
- With AI, this becomes a community resource for building models

https://mdrepo.org

Attribution is a problem

- Spurious correlations unlikely to be a problem
- Context are easy to establish, reducing the potential for bias
- Attribution – assessing the value of a contribution **is difficult**
  - Immediate value to science *versus*
  - Value of contribution to model development
- Other issues:
  - Distributed repositories and distributed (not federated) learning
  - Model bias introduced by (near) duplication of simulations

Thanks

- Duke Pauli: Agroecosystem Research in the Desert (ARID)
- Matt Rahr: Cyber Experiment Station
- Nirav Merchant: Data Science Institute
- Connor Osgood: YCEDA AgTech Program
- Jeff Larsen: Biosphere 2
- Ash Black: InfoSci
- Travis Wheeler: MDRepo & Pharmacy
- David Ebert: Chief AI and Data Science Officer

THE UNIVERSITY OF ARIZONA
College of Information Science

It’s the data (lifecycle), stupid!

THE UNIVERSITY OF ARIZONA
College of Information Science