Toward a Health Data Commons for Arizona

Arthur "Barney" Maccabe
Executive Director
Institute for Computation and Data-Enabled Insight
The University of Arizona

Presentation for SOS27
Engelberg CH
March 20, 2025

RESEARCH, INNOVATION & IMPACT
Institute for Computation & Data-Enabled Insight

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
Institute for Computation & Data-Enabled Insight

INSPIRATIONAL WORDS

“

This is an opportunity to add a quote that supports your presentation.

NAME

THE UNIVERSITY OF ARIZONA

INSPIRATIONAL WORDS

“

Data is the new bacon

Doug Kothe

THE UNIVERSITY OF ARIZONA

Background

FROM THE FIRST DAY:

AI = Data + Energy (compute)

Stargate

- Announced January 21, 2025
  - LLC consisting of OpenAI, SoftBank, Oracle, and MGX (investment firm)
- $500B investment in AI infrastructure in the US by 2029
  - $100B in 2025
  - “Manhattan Project” for AI (Industry exclusive)
- Expected federal role – remove regulatory barriers
  - Energy
  - Data
- Expected academic/laboratory role – who knows?

FROM THE SECOND DAY:

Sustaining Software (products) is Hard

Data is harder…

Strategy

Get Data

Academics will not have significant assets in energy (and computing)

We probably don’t even have much data

Focus on state data

Goal: Establish the ABOR universities as ***trusted stewards*** of state data

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

Arizona Research Computing (ARC)

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

- Arizona Board of Regents (ABOR) oversees the three public universities in Arizona
- ARC includes the three universities and Sun Corridor Network

ARIZONA BOARD OF REGENTS
ASU ◆ NAU ◆ UA

ASU NAU

Advancing Arizona's research landscape through cutting-edge computational infrastructure and data stewardship, empowering innovation and community growth.

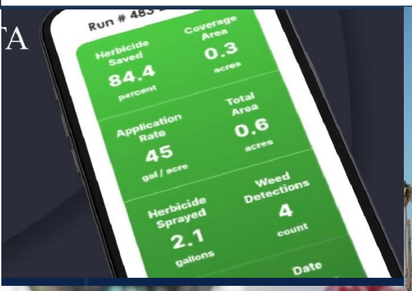
ARIZONA RESEARCH COMPUTING

https://suncorridor.org/research-support/arizona-research-computing/

An Example

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.

Connecting fields in Yuma County

Image courtesy of Connor Osgood

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

RESEARCH, INNOVATION & IMPACT
Institute for Computation & Data-Enabled Insight

Health Data Commons: The Pitch

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

Three Whys

Why Health Care?

- Arizona faces significant public health challenges: rural health disparities, rising chronic conditions, and resource constraints.
- Improve health and drive impactful research outcomes.

Why Data Commons?

- Data silos limit collaboration and coordination.
- Provide a secure enclave for integrating and analyzing data assets, enabling actionable health data statewide.

Why Now?

- Rapidly emerging technologies that allow deeper integration.
- Statewide Data Modernization Initiative (DMI).
- Momentum from tri-university collaboration.

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

Vision

- A state-of-the-art data platform integrating health data from multiple sources, enabling AI-driven insights.
- Supporting better health forecasting, optimized resource allocation, and improved healthcare outcomes.
- Co-designed by Arizona's public universities, Arizona's health agencies, HIE, counties, and other stakeholder organizations.
- Providing visibility into priorities across involved organizations in support of shared governance and collaboration.

THE UNIVERSITY OF ARIZONA

Process

Engagement, Engagement, Engagement

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

Pitch

Present to Arizona Data Modernization Initiative

January 13, 2025

Requirements

Workshop to define use cases

April 10, 2025

Governance

Workshop to develop governance model

Summer 2025

Data Management

A Data Mesh Model

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

Federation of data sources to maintain sovereignty and flexibility.

Key components:

- Data ingestion: Real-time integration from diverse sources (e.g., ADHS, AHCCCS, tribal data, clinics).
- Analytics: AI/ML for predictive insights.
- Security: Robust data privacy and ethical standards.
- Visualization: Intuitive analytical tools for policymakers and researchers.

Governance: A tri-university-led initiative with multi-stakeholder oversight and collaboration.

Open Data Access

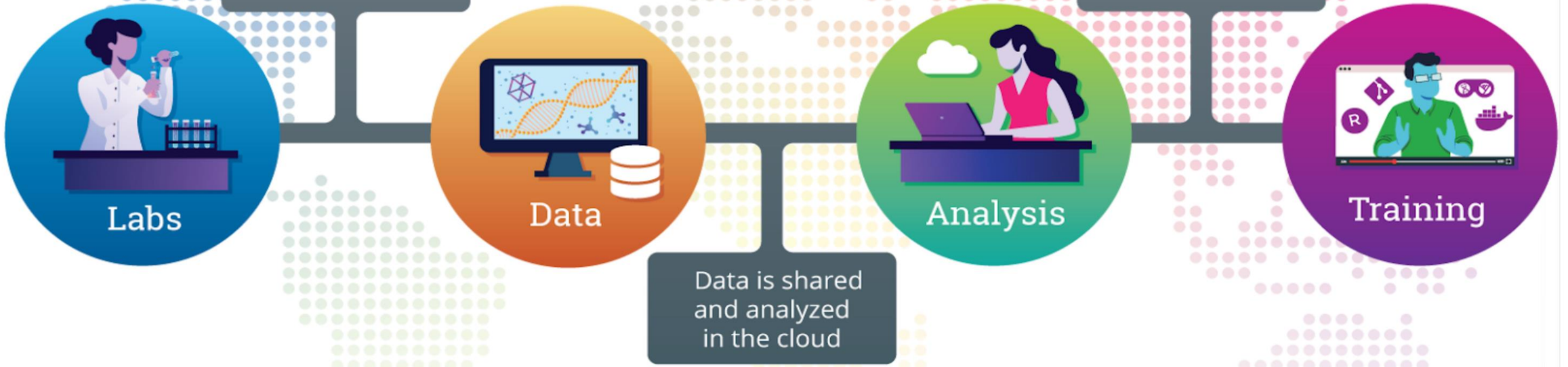
- Emphasis on the public good and open science
- Open as possible, closed as necessary
  - Respect regulations, e.g., HIPAA (Health Insurance Portability and Accountability Act)
  - Respect data sovereignty, data owner can impose additional restrictions/requirements
- Public access will likely be based on the US Census Data Access and Dissemination System (DADS), including
  - Pre-built data products
  - Differential privacy

Restricted Access to Microdata

- Users with specific needs are admitted to the “walled garden”
  - Background checks; binding agreement with consequences for disclosure
- Common, “all or nothing” access is not adequate
  - Insider threat; data sovereignty
- Bind role to functional access, e.g.:
  - Health care provider can access an individual’s EHR (while treating the individual), but not their financial information.
  - Insurance provide can see procedures performed, but not physician notes.
- Functions defined by data owner (respects data sovereignty)
- Function-based Access Control (based on functional encryption)

CYVERSE® Public

100,000 users | **1500 publications** | **11 PBs of data** | **40,000 trained**



CYVERSE® Professional

- Federation with local and commercial cloud and high-performance computing
- Integration with local user identity management systems
- Security compliance

CYVERSE® Health

- Secure Perimeter (VPN)
- HIPAA Compliant Platforms inside of Perimeter
- XNAT
- Training

CYVERSE® Defense

- ITAR Compliance
- Receive data from multiple sensors
- Policy based data visibility and sharing
- Support for multiple teams and data partitioning

CYVERSE®

THE UNIVERSITY OF ARIZONA
Institute for Computation & Data-Enabled Insight

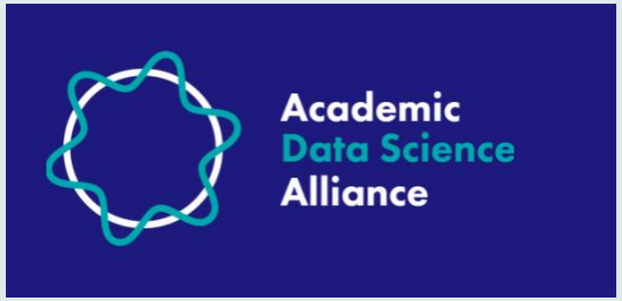
- History: iPlant Collaborative -- data sharing for plant science
- Components
  - Based on iRODS
  - Data store (establishes ownership)
  - Discovery environment (registration and sharing)
- Basic strategy:
  - Bind data (reference) and functionality into a container
  - Share container (not raw data) with appropriate role(s)
- Critical extension
  - Current strategy: Data moves to compute
  - Move code to data, or extend trusted compute enclave

THE UNSPOKEN PROBLEM

Business/Sustainability Model

Thanks

Ask me about ADSA Membership



https://academicdatascience.org/get-involved/adsa-membership/

THE UNIVERSITY OF ARIZONA
Institute for Computation
& Data-Enabled Insight