National Center for Supercomputing Applications

Facts, Resources & Collaborations

Daniel S. Katz
Chief Scientist, NCSA
Associate Research Professor, CS, ECE, iSchool

5 April 2024
CSBS & NCSA Data Science Kickoff

ILLINOIS
NCSA | National Center for Supercomputing Applications

The university

- Mission: The University of Illinois at Urbana-Champaign is charged by our state to enhance the lives of citizens in Illinois, across the nation and around the world through our **leadership in** learning, **discovery**, engagement and economic development.

- Research: At Illinois, our focus on research shapes our identity, permeates our classrooms and fuels our outreach. **Fostering discovery and innovation is our fundamental mission.** As a public, land-grant university, we have the responsibility to create new knowledge and new ideas and translate these into better ways of working, living and learning for our state, nation and world.

NCSA

- Began as one of 5 NSF-funded centers for supercomputing
  - The NSF supercomputing program emerged from a proposal from UIUC!
- Long history; international reputation; a leader in computing
- Today
  - Large and varied research portfolio (over $50M annually in sponsored research)
    - Computing and Data essential to many areas of scholarship
    - Ranges from software and applications development to designing and operating supercomputers
  - Partners across campus in interdisciplinary research projects
  - Operates a state-of-the-art data center and provides research computing to campus and the nation
  - 244 staff & postdocs; 187 faculty and center affiliates, from 58 departments, 9 colleges/schools; 206 students

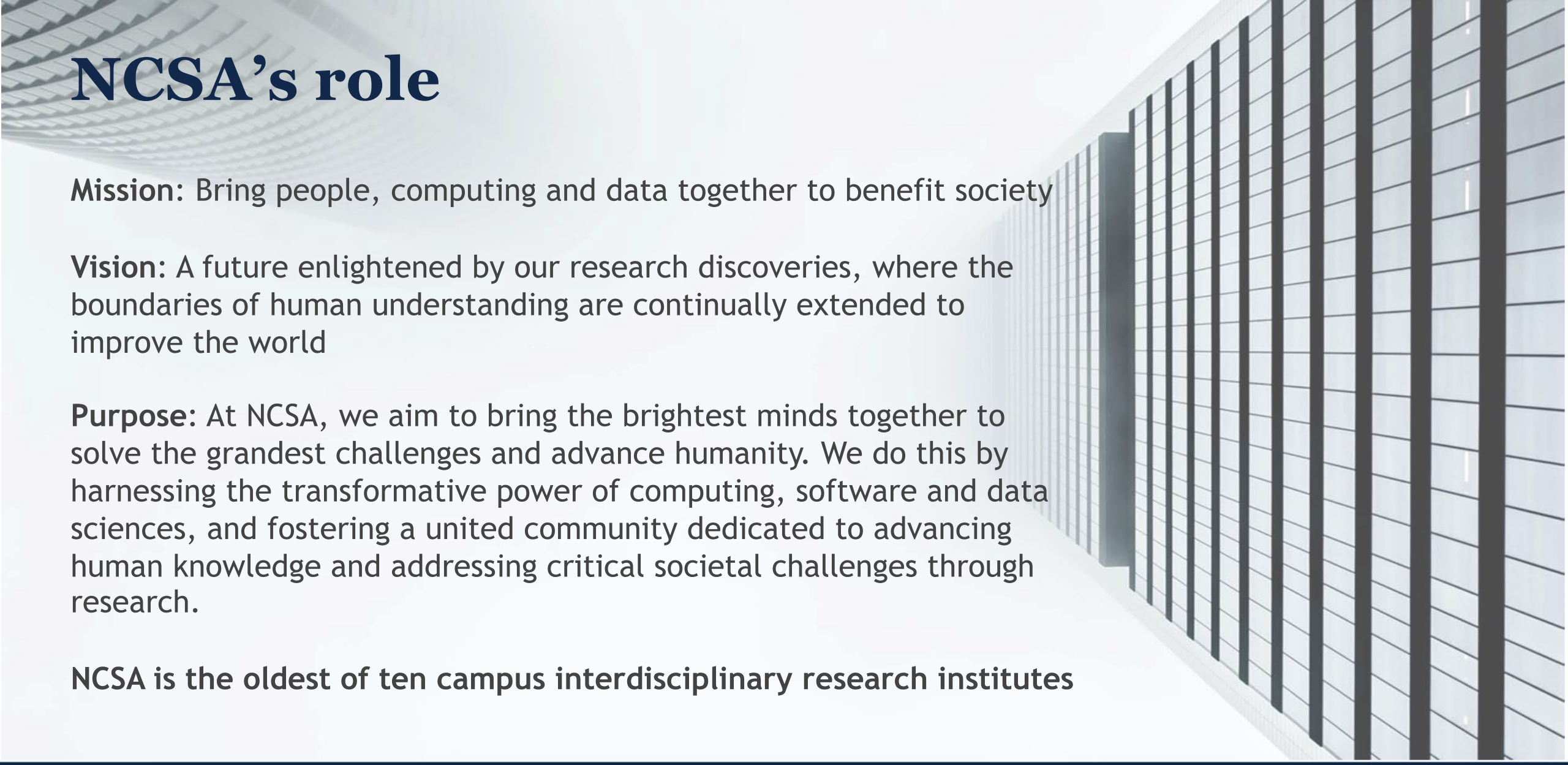
NCSA's role

Mission: Bring people, computing and data together to benefit society

Vision: A future enlightened by our research discoveries, where the boundaries of human understanding are continually extended to improve the world

Purpose: At NCSA, we aim to bring the brightest minds together to solve the grandest challenges and advance humanity. We do this by harnessing the transformative power of computing, software and data sciences, and fostering a united community dedicated to advancing human knowledge and addressing critical societal challenges through research.

NCSA is the oldest of ten campus interdisciplinary research institutes



Leadership

Systems & Expertise

- **Computing:** HPC, HPC-AI, Storage, Extreme Scale Machine Learning, HTC, HPC for sensitive data
- **Expertise:** ACCESS (national), state, system, campus

Technologies

- AI & Machine Learning
- Data Science
- Software

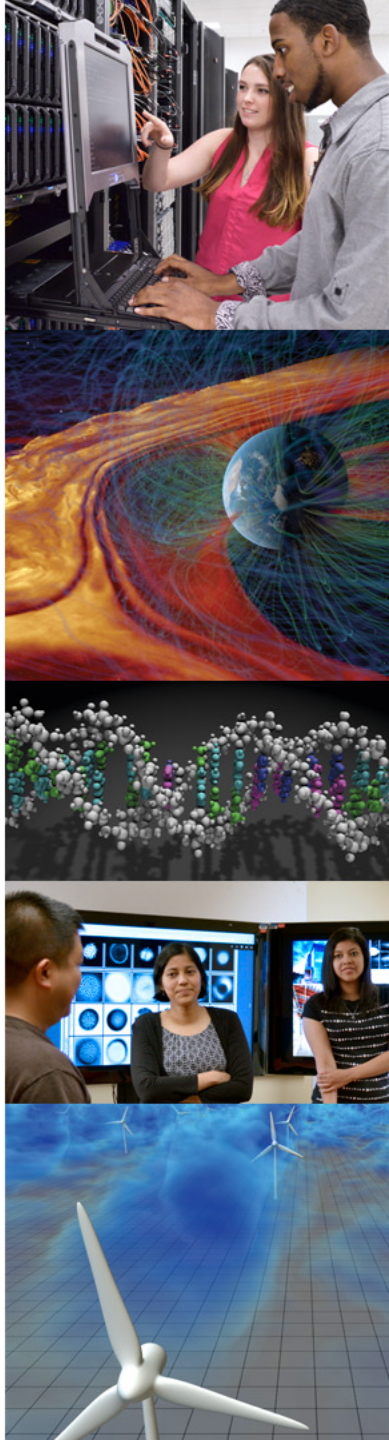
Disciplines

- Artificial Intelligence
- Astronomy
- Arts & Humanity
- Health Sciences
- Digital Agriculture
- Earth & Environment
- Quantum
- Engineering & design

NCSA Industry

- Health
- Manufacturing
- Technology

Resources



ILLINOIS NCSA

NCSA technologies & expertise

Systems and Facilities

- Delta (HPC)
- DeltaAI (HPC & AI, coming in summer)
- HAL (ML)
- Illinois HTC program
- Radiant (cloud)
- Nightingale (HIPPA-compliant)
- NPCF Facilities
- Illinois Computes Research Notebooks (ICRN)

Data and storage

- Integrated Data and Database Services
- Storage Enabling Technologies
- Taiga & Granite

Networking & security

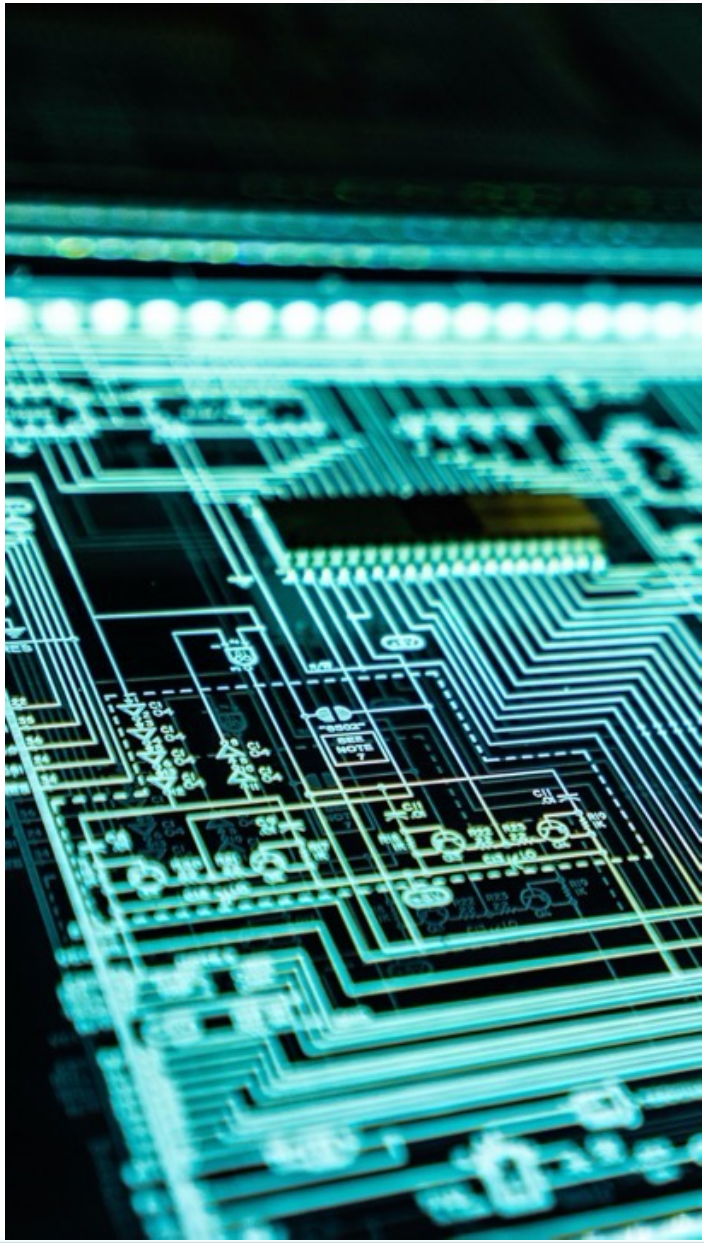
- Cybersecurity
- Network Engineering and Research

Software

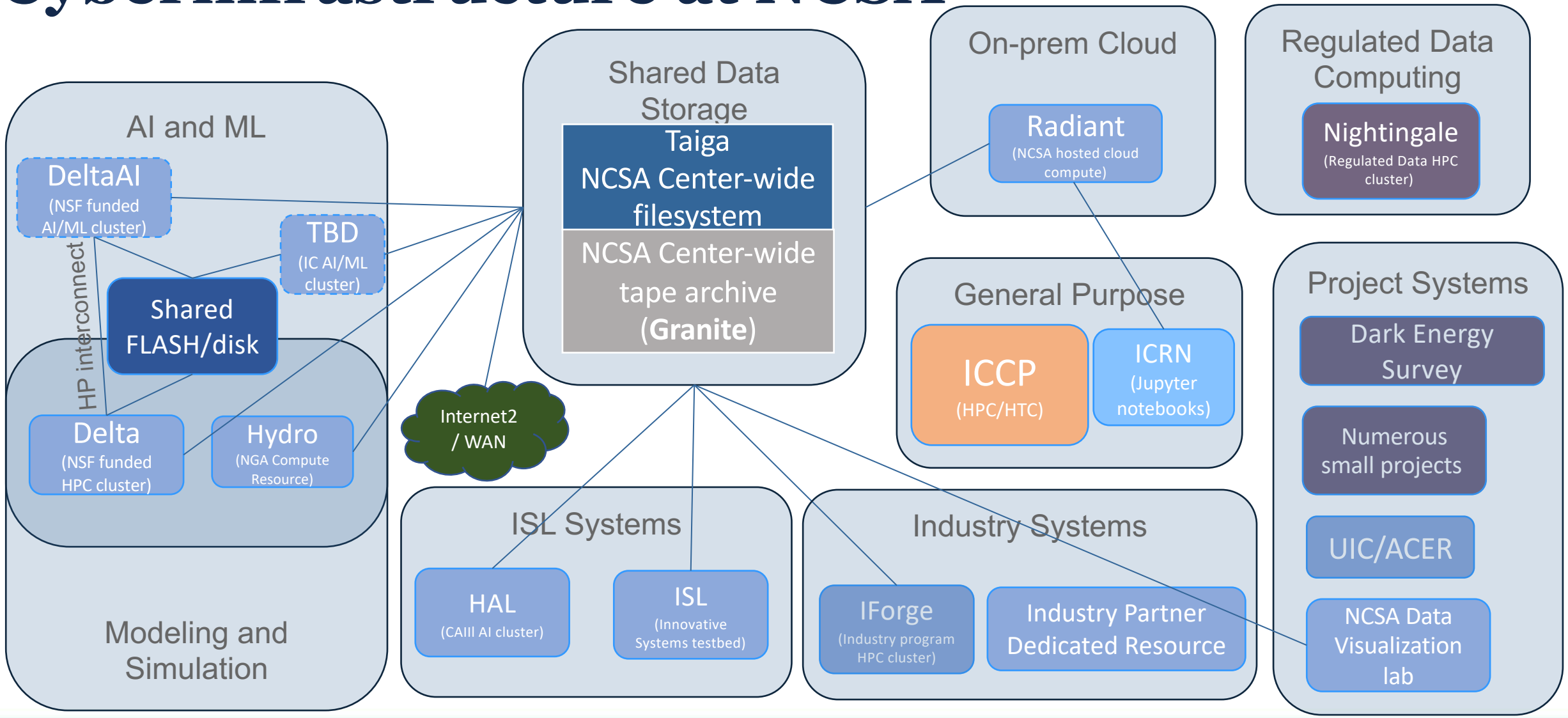
- Innovative Software and Data Analysis
- Visual Intelligence Group

Services

- Innovative Technology Services
- Scientific Computing Services & System Administration
- Advanced Application Support
- Scientific Engineering Application Support
- Technology Management Group



Cyberinfrastructure at NCSA



Compute and data resources at NCSA

Name	Capabilities	Use Model	Details
Delta	GPU-rich supercomputer for simulation, AI/ML, Data Science (NSF funded)	Batch/Browser	440 A100, 400 A40, 8 MI100, 124 dual Milan (NSF, 10% for UIUC)
DeltaAI	GPU-rich supercomputer for AI/ML (NSF funded)	Batch/Interactive	320 GH200 (NSF, 10% for UIUC)
TBD	GPU-rich supercomputer for AI/ML	Batch/Interactive	80 GH200
Nightingale	Mixed CPU/GPU, HIPAA compliant, ePHI/CUI secure	Batch/Interactive	Suitable for HIPAA, CUI, FERPA
Radiant	On-Prem Cloud – OpenStack	Interactive	31,000 vcpus, 8 vgpus
Hydro	Mixed CPU/GPU for MPI/AI/ML	Batch	55 nodes incl 18 A100
HAL	GPU-rich cluster for AI/ML	Batch/Interactive	16 P9 with 64V100; DGX2; FPGA
Taiga/Granite	Data Storage Services – Disk and tape	Shared File System	18PB disk 40PB tape; can grow
Campus Cluster	Mixed CPU/GPU “condo style” cluster	Batch	769 CPU, 50 GPU nodes

Details and up-to-date config: https://wiki.ncsa.illinois.edu/display/USSPPRT/NCSA+Allocations
DeltaAI and "TBD" to be delivered April-May 2024, in operation Summer 2024

Delta

- **Targets the computational resource needs of simulation, data science, and AI/ML**
- **90% allocated nationally via ACCESS, 10% for Illinois strategic priorities (e.g., Center proposals, faculty recruitment/retention, research initiatives)**

- Balances CPUs and GPUs for both traditional and newer workloads and user practices, using both together, and migrating between
- Includes large memory for in-memory databases
- Supports access via command line, Open OnDemand, Science Gateways, DTNs, etc.
- Allocations via ACCESS can be fast and easy, proposal requirements tied to request sizes

- NSF's largest GPU resource currently
- 124 AMD CPU nodes, 200 NVIDIA quad GPU nodes, 5 8-way A100 GPU nodes, 1 AMD MI100 CPU/GPU big memory node
- Supported by 7 PB disk and 3 PB flash
- Highly scalable interconnect fabric

NVIDIA ddn Hewlett Packard Enterprise

DeltaAI Bridging the gap between AI/ML computing demands and today's capabilities

- **Targets the computational resource needs of AI/ML workloads**
- **To be deployed in 2024**
- **90% allocated nationally via ACCESS, 10% for Illinois strategic priorities (e.g., Center proposals, faculty recruitment/retention, research initiatives)**

- Working with AI/ML projects to define and enable new & existing access & usage modalities best suited to the AI/ML community: Web interfaces, interactive use, persistent services and node use, etc.
- Will build upon existing Delta infrastructure, such as Open OnDemand, Science Gateways, DTNs, etc.
- Working with NSF, ACCESS and the AI/ML community to improve the allocation process for AI/ML projects

Will double Delta's performance, with:

- 21 FP64 PF and 633 FP16 PF from 320 NVIDIA Grace Hopper GPUs with 96 GB per GPU (384 GB GPU mem per node)
- Supported by 14 PB of storage at up to 1 TB/sec
- Highly scalable interconnect fabric

NVIDIA ddn Hewlett Packard Enterprise

Software

- Since its creation in 1986, NCSA has led in working with & supporting scholarly communities through the development, deployment, and use of new computing and software technologies
  - From NCSA Telnet, NCSA Mosiac, …, to dozens of current projects
  - Work with individual projects to apply existing technologies
    - E.g., IN-CORE: hazard modeling and management service
  - Work with multiple projects with common needs to build new technologies
    - E.g., Clowder: web-based content repository, providing a platform automatic metadata extraction, automatic data previews, distributed/heterogeneous data sources, social curation, provenance, and data publication
  - Across disciplines, across languages, across platforms
  - Leading the national research software engineering movement, to provide stable career paths to research software developers (via US Research Software Engineer Association)
    - Now 50+ RSEs in NCSA's software directorate, and 70+ across NCSA

Research Consulting offers a unique service to HPC users

A comprehensive approach to provide end-to-end guidance and expertise for research to all NCSA users and partners, with:

- Providing 24x7 helpdesk support, user training, and documentation
- Highly skilled staff with advanced degrees and in-depth experience
  - In fields such as computer science, physics, engineering, genomics, crop science, mathematics, statistics, astrophysics, cosmology, astronomy, earth sciences, atmospheric and climate science
- Utilizing skills, programming, and technologies
  - Such as multiphysics numerical analysis, machine/deep learning with data analytics, bioinformatics, linear algebraic methods, molecular dynamics, statistics, workflows, high performance computing, high throughput computing, GPU systems, advanced visualization techniques
- To work closely with research teams and advance scientific discovery

Illinois Computes!

- New university & system commitment to transform scholarship by making computing available to all researchers
  - For any researcher to get started using computing & data without having to first find funding
  - Reduce “friction” in applying computing and data to research
  - Broadening access, use of computing & data in research across the entire UIUC campus & UI system
  - Complements other efforts on campus and system, including ResearchIT and efforts at college, institute, and unit level
- Computing systems (campus & system: $35m) & research consulting (campus: $15m)
  - 16 new nodes providing first-ever “campus” queue for campus cluster
  - Systems includes AI/ML support, data storage (5 TB/researcher) near compute, secure computing (e.g., for HIPAA, CUI, FERPA data)
  - Significant AI system (20 nodes of 4x Grace-Hopper), to be deployed at the same time as DeltaAI
  - Jupyter notebooks for interactive computing (ICRN)
  - Consulting: human services to help researchers use systems effectively
- Established Campus Advisory Committee

Illinois Computes Research Notebooks (ICRN)

- Free Jupyter notebooks for researchers
- Key Benefits:
  - Open to All Researchers: Any researcher in need of a basic computational resource
  - User-Friendly: Intuitive Jupyter notebook environment
  - Go remote: Provides sufficient resources to replace local code execution
  - Instant Access: Immediate access upon login
  - Resources: up to 4 CPU cores, up to 8 GB RAM, 100 GB of persistent storage (ICCP filesystem), GPUs (coming soon)

Get access now!

https://go.ncsa.illinois.edu/jupyter

Illinois Computes services

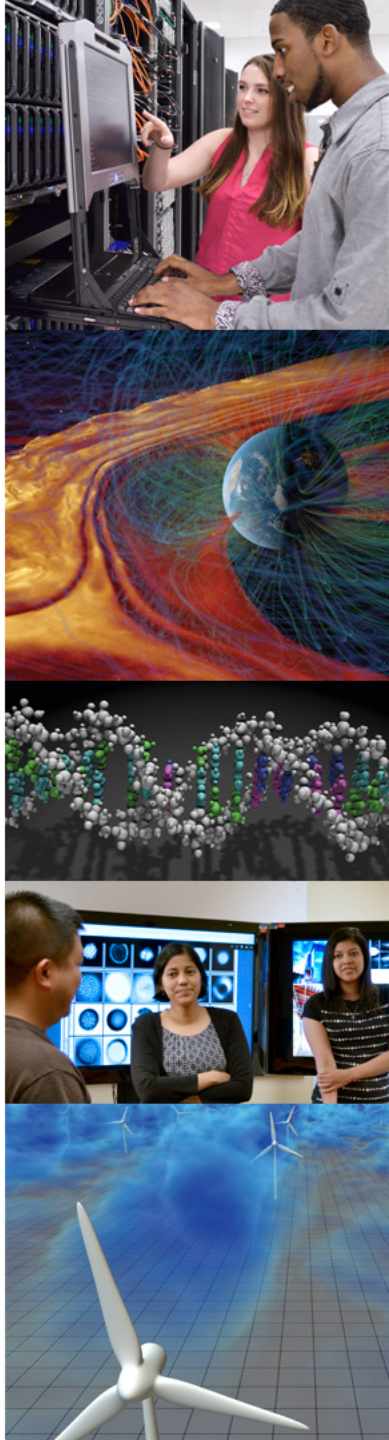
(as of mid-March, 2024)

236	New projects on computing or data systems
269	Support requests
44+	Departments or research units served
50+	Referrals to Delta through ACCESS
110	Ongoing collaborations
10	Out of 14 colleges
31	Live training events/yr
1452	Registrants
18	Asynchronous training topics
548	Enrollees

How to request resources

- Delta and DeltaAI (Illinois Allocation)
  - https://delta.ncsa.illinois.edu/
  - https://xras-submit.ncsa.illinois.edu/login (requires NCSA identity)
  - DeltaAI Information will be available after DeltaAI is installed
  - NCSA is part of the National AI Research Resource (NAIIR) Pilot:
    - https://nairrpilot.org/allocations
- NSF ACCESS
  - https://access-ci.org/
  - Including use of Delta (and DeltaAI when it becomes available)
- Illinois Computes
  - https://computes.illinois.edu/
  - Request resources or help/support at https://computes.illinois.edu/submit-a-request/
    - How do I get started?
    - How do I optimize my code?
    - I need help visualizing my data

Collaborations



ILLINOIS NCSA

Centers & campus project collaborations

Center for Digital Ag (CDA) & AIFARMS & CROPPS	Delta & DeltaAI	NSF Materials Science NRT (DIGI-MAT)	C3.ai Digital Transformation Institute	Center for Astrophysical Surveys (CAPS)
Center for Artificial Intelligence Innovation (CAII)	Midwest Big Data Hub (MBDH)	Illinois Mayo Alliance	Center for Exascale-Enabled Scramjet Design (CEESD)	Healthcare Innovations Program Office

Working with NCSA: People

For faculty & staff
- NCSA Fellows (not in 2025)
- Affiliates (faculty & staff)
- Fiddler Fellowships

For students
- SPIN (Students Pushing Innovation)
- ML REU (FoDOMMaT)
- Digital Ag REU (CDA REU)
- NCSA International Research Internship
- NCSA Frankel Uni High Scholar Program
- Fiddler Fellowships
- Cyberinfrastructure Professional Intern Program (CIP)

Centers
- Center for AI Innovation
- Center for Digital Agriculture
- Healthcare Innovations Program Office (HIPO)
- Midwest Big Data Hub

Labs:
- Advanced Visualization Lab
- Data Analysis & Visualization Lab
- Data Exploration Lab
- vi-bio lab
- Innovative Systems Lab (ISL)

Programs
- Computational Science & Engineering
- NCSA Industry

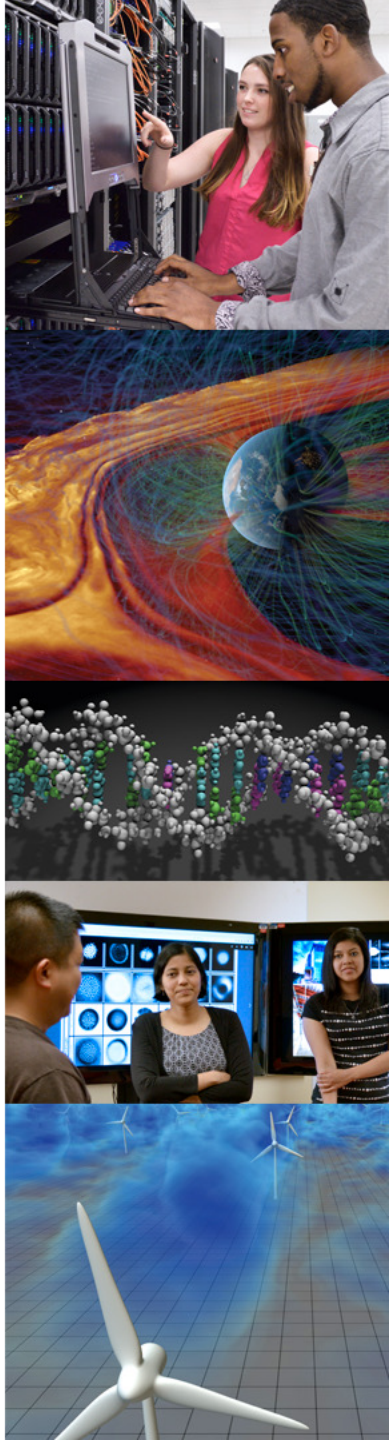
2nd Annual NCSA Student Research Conference

- **April 23, 3:00-6:30 PM, NCSA building**
  - Register by April 15, 2024 at https://forms.illinois.edu/sec/785875075

Schedule:

3:00 - 3:05 p.m. | Welcome & Opening Remarks by Bill Gropp, NCSA Director - NCSA Atrium
3:05 - 4:00 p.m. | Poster Session - NCSA Atrium
4:00 - 4:10 p.m. | Break - NCSA Atrium
4:10 - 4:55 p.m. | Oral Presentations - NCSA 1030 & 1040
4:55 - 5:00 p.m. | Break - NCSA Atrium
5:00 - 5:45 p.m. | Mixed Panel - NCSA Auditorium
5:45 - 6:30 p.m. | Industry Interactions & Dinner - NCSA Atrium
4:00 - 6:00 p.m. | Workshop - NCSA 1104

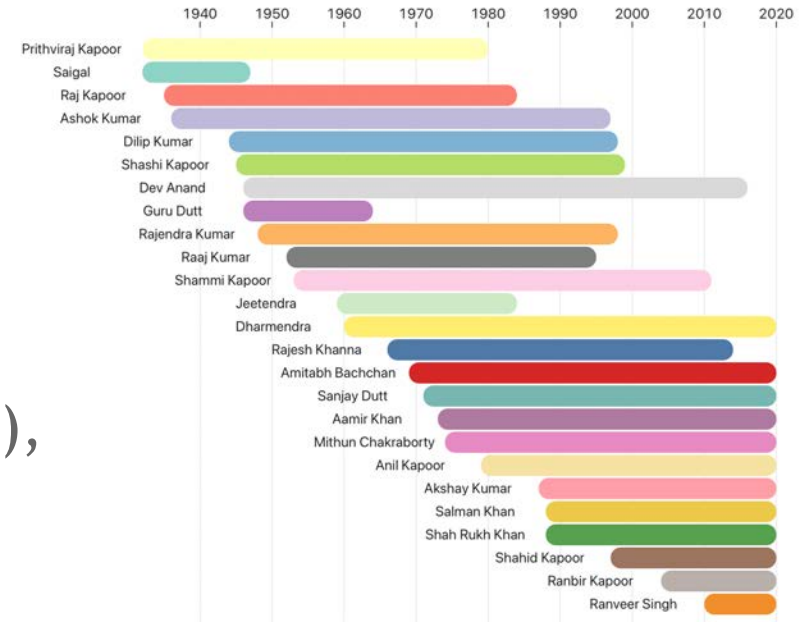
Example Projects



ILLINOIS NCSA

Indian Cinema in Context: Interactive Film History Archive and Tools

- Team: Rini Mehta, Comparative and World Literature; Rob Kooper, Sandeep Puthanveetil Satheesan, Luigi Marini, Kalina Borkiewicz, NCSA; Donna Cox, NCSA/Art & Design
- Fellow (2018-2020) goal: Collect data to analyze Indian cinema from the perspective of the viewer (what was shown where and when), to understand the text (understanding presence, e.g., speaking, screen time, of characters), and to understand relationships (family connections)
- Has led to overall interest in using software to connect film and literature data
- NCSA part: data and software expertise, conceptualization and information about various tools for network graphs and visualizations
- Results: LAS Fellow in a Second Discipline (Computer Science), UI Presidential Grant, continued collaboration proposals, summer workshop with NCSA



Data science for the social good: Enhancing humanitarian efforts for forcibly displaced populations

- Team: Angela Lyons, Agricultural and Consumer Economics & NCSA; Aiman Soliman, NCSA & Urban and Regional Planning; plus more, see Angela's earlier presentation
- Goal: assist in formulating more targeted and effective policy recommendations for cash assistance programs for forcibly displaced persons, leading to reports, policy documents, machine learning models, understanding of key features
- NCSA part: computing, data, staff, and student resources; project and tool organization
- Results
  - Team chose to work with NCSA because
    - Resources, people, programs to enable success
    - Collaborative team working for a common purpose
  - NCSA learned
    - Addressing data challenges, including collection, storage & security, long-term preservation
    - Creating the right atmosphere & culture, and doing upfront work on agreements

CPRD Medicaid waiver evaluation

- Team: CPRD & NCSA (Research Data Engineering and Scientific & Engineering Applications Support (SEAS) Groups)
- Goal: Evaluate the results of a specific Medicaid waiver related to substance use disorder treatment
- NCSA part: database hosting services, exploratory data analysis, data transformations, analysis tools, training
- Results
  - Team chose to work with NCSA because
    - It's a collaborator, not just a provider of storage/compute resources
    - NCSA focused on ensuring HIPAA-compliant, monitored, secure data, letting CPRD team focus on their unique challenges
  - NCSA learned
    - Data issues can vary widely with field: small/large, simple/complex, analysis tools
    - Work with a team can develop shared expertise, leading to new ways of thinking about problems and potential follow-ons

Partner with NCSA on your next research challenge!

- Expert technical staff
- Project managers, program & grant administration
- Computing, software, data resources
- Data analytics & visualization
- Educational and DEI programs