

University of Arizona Cyberinfrastructure and Research Support 2022

Introduction

The University of Arizona (UA/UAritona) is the Arizona Land Grant University and a research-intensive University. As a Land Grant Institution, we are responsible for the Arizona [Cooperative Extension](#) System present in every county and directly charged with enhancing statewide economic development. It is critical for UA to provide a comprehensive Cyberinfrastructure (CI) plan and roadmap that facilitates collaboration and enables faculty, staff and students to conduct leading-edge research. The UA CI plan aligns with and takes inspiration from the 2007 NSF Cyberinfrastructure Vision for 21st Century Discovery report. As defined in the report, CI consists of computing systems, data storage systems, data repositories, specialized instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable breakthroughs that would otherwise not be possible. UA cyberinfrastructure, policies, and training create an ecosystem of services to support research throughout its lifecycle.

The UA CI plan has been updated and refocused to align with the [2017 NSF 10 Big Ideas](#), which provide a vision for research in 2026. The following committees provide research focus and strategy for implementing CI at UA: Research Computing Governance Committee (RCGC), and the Data Science Research and Training (DSRT) committee. RCGC has subcommittees tasked with leading strategic advisory efforts: High Performance Computing (HPC) Technology Refresh, Data Management & Curation, and HPC Policies & Procedures. Subcommittee membership is intended to be transdisciplinary.

University of Arizona Research Cyberinfrastructure Goals

As research becomes increasingly complex, it requires more computational resources and associated training to enable researchers. The primary objective of the CI plan is to provide details on the necessary infrastructure, tools, technologies, training, and supporting expertise that will ensure that our researchers are able to accomplish their research goals and successfully collaborate with colleagues around the globe. Delivering the promise of modern research and digital literacy presents challenges for existing computing, network infrastructure, storage, and software tools. The goals outlined below are key to providing UA researchers with supported and sustainable cyberinfrastructure to meet current and future research needs.

1. UA Research Cyberinfrastructure Ecosystem

The computing, networking, and storage demands of campus research are continuously expanding in scope and complexity. This presents a challenge to both central and discipline-specific infrastructure and support staff. Providing a more robust ecosystem of infrastructure, integrating central and distributed resources, and providing a coherent campus-wide cyberinfrastructure addresses these challenges. The CI ecosystem is developed and sustained through agreement and cooperation among myriad university organizations.

1. UA Research Cyberinfrastructure Ecosystem	1
1.1 Research Technologies Cyberinfrastructure Resources	3

1.2. Storage	4
1.3. Research Data Center Facilities	4
1.4 Regulated Research Data and Computing Environments	5
Legal Aspects of Regulated and Controlled Data	5
Regulated Research Environments (CUI, HIPAA, etc)	5
1.5 CyVerse Computational Resources	6
1.6 XSEDE and Open Science Grid (OSG) National Computing Resources	7
1.7 UITS Network Capacity	7
1.8 Data Science Institute	8
1.8.1 Project-based research support	9
1.8.2 Data Science Training	9
1.8.3 KMAP Knowledge Map	10
1.8.4 Course Catalog Search	10
1.8.5 Center of Excellence in Data for Society (CEDS)	10
1.9 TRIPODS Initiative	11
1.10 Research, Innovation & Impact	12
1.11 The Office of Digital Innovation & Stewardship (ODIS)	13
1.11.1 Computational Research Analysis and Methods (CRAM)	13
1.11.1.1 Computational Research Support Services	13
1.11.1.2 Research Data Management	13
1.11.1.3 Research Data Repository	14
1.11.2 Scholarly Communication and Research Information Management (SCRIM)	14
1.11.2.1 Scholarly Communication Services	14
1.11.2.2 Campus Repository	15
1.12 The CATalyst Studios	15
1.12.1 Data Studio	15
1.13 Advanced Visualization at the University of Arizona	15
1.14 Health Sciences	16
1.14.1 Surveys and Clinical Databases	16
1.15 BIO5	16
2. Improving Institutional Efficiency	17
3. Developing Software Tools	17
4. CI Support and Training	18
4.1 Data Science Resources & Training	18
4.1.1 Data Science Ambassadors	19
4.1.2 Data Science Fellows	20
4.2 Research Bazaar Arizona (ResBazArizona)	20
5. Support for NSF Ten Big Ideas	21
5.1 Future of Work at the Human-Technology Frontier	21
5.2 Growing Convergence Research	21
5.3 Harnessing the Data Revolution	21
5.4 Mid-scale Research Infrastructure	24
5.5 Navigating the New Arctic	24

5.6 NSF 2026	25
5.7 NSF INCLUDES	27
5.8 Quantum Leap	28
5.9 Understanding the Rules of Life	28
5.10 Windows on the Universe	29
6. Acknowledgements	30

1.1 Research Technologies Cyberinfrastructure Resources

Last Updated: 2022-12-15

Point of Contact(s): Chris Reidy, chrisreidy@arizona.edu

The [University Information Technology Services \(UITS\)](#) team has maintained centrally funded and managed research computing infrastructure, infrastructure administration, and associated research consulting for High Performance Computing (HPC), Visualization, and Statistics since 1985. Recurring funding from UA has allowed for refresh of the centrally supported HPC systems every 3-5 years. The HPC resources are available, at no cost, to all campus researchers.

UA researchers have an option to fund additional HPC compute nodes and leverage the central infrastructure and system administration support, known as 'buy-in.' This allows researchers to obtain high-priority access to the purchased nodes, which are centrally managed as part of the larger system. Unused compute cycles that might be available from the buy-ins are available to other campus researchers, at a lower priority. Buy-in can improve grant competitiveness due to lower costs that must be included in the grant (i.e., hardware only, no operations costs); evidence of campus cost-sharing; and more positive funding agency review of cost-effective centrally administered facilities. The following systems are available at no cost to researchers:

- El Gato general purpose cluster purchased 2013 from NSF MRI grant 1228509 awarded to faculty in Astronomy and the School of Information and central UA funds with 2176 Intel cores, FDR IB interconnect, 26TB of RAM, 140 NVIDIA Tesla K20X GPUs.
- Ocelote general purpose cluster (purchased 2016) with 11528 cores, 57TB of memory, FDR Infiniband (IB) Interconnect, 10G ethernet, and virtual shared memory (vSMP) software. Additionally, a 48 core dedicated high memory node with 2 TB RAM and 48 NVIDIA Tesla P100 GPUs are available.
- Puma general purpose cluster purchased 2020. The vendor is Penguin Computing. The system specifications are:
 - 48 Altus XE2242 CPU Nodes with 96 CPU x 4 = 384 CPU, 2.4GHz, 512GB RAM per node
 - 6 Altus XE2214GT GPU Nodes with 96 CPU, 2.4GHz, 512GB RAM, 4 32GB V100S per node
 - 2 Altus XE1212 High Memory Nodes with 96 CPU, 2.4GHz, 3072GB RAM per node
 - Total CPUs: 19,200 2.4 GHz AMD (EPYC 7642 Rome)
- Since the original purchase of Puma, researchers have added capacity with their funding that has taken the core count to 28952, which is a 50% increase.

Access to UA's 2020 Puma cluster includes 100,000 CPU hours/month, the 2016 Ocelote HPC system includes 70,000 CPU hours/month and 7000 CPU hours/month on El Gato per PI group, with additional windfall processing time available on these systems.

UITS has six full-time consultants with expertise in HPC, statistics, and data visualization; and four full-time staff dedicated to UA HPC Systems Administration. Research Computing Support consulting includes multi-core/processor programming, support for scaling and benchmarking of parallel code, Singularity container technologies, and expertise in specialized accelerator technologies such as NVIDIA general purpose graphical processing units (GPGPU or GPU).

1.2. Storage

Last Updated: 2022-12-14

Point of Contact(s): Chris Reidy, chrisreidy@arizona.edu

Centrally administered HPC storage follows the same funding, support and upgrade models as the HPC computational resources. The UA HPC storage array is housed in the same research data center as the HPC.

The Qumulo P184 SSD based storage array was designed for reliability, performance, and scalability. To accommodate continued I/O demands in research data, this array provides 2PB of raw capacity, 65GB/s of read bandwidth and 44GB/s of write bandwidth. This is an all-flash solution which is very fast and very expensive. One of the benefits to researchers is that the computation times are not delayed by highly intensive IO operations from multiple users.

This storage array is mounted on all three clusters and contains all of the software packages which are installed as modules, which means that the exact same applications can be run on any cluster.

Access to the HPC resources includes 50GB of home directory space, 500GB of space for each PI and an allocation of time-limited space up to 20TB available on request to PI's.

In late 2020 another array was provisioned that is available for rental. The rate is \$47.25 per TB per year. It is mounted as /rental but not directly to the HPC clusters.

An AWS solution is available that is more like archival. It uses AWS Intelligent Tiering between S3 and Glacier and is subsidized by UITS.

1.3. Research Data Center Facilities

Last Updated: 2022-12-14

Point of Contact(s): Chris Reidy, chrisreidy@arizona.edu

UA has two facilities to house research computing resources:

1. Research Data Center (RDC): 1200 ft² raised floor data center designed for water-cooled racks dedicated to centrally managed research computing systems and large grant funded systems, with capacity for 40 standard racks. RDC capacity allows UITS to maintain four generations of centralized large computer clusters simultaneously, providing continuity of research while each new system is phased in.
2. Co-location Data Center: 1900 ft² of raised floor data center space for air-cooled research computing systems purchased and administered by colleges, departments, and research projects, with capacity for 70 standard racks.

Other than a nominal installation cost, no bandwidth or other recurring charges are levied for co-located research systems in these facilities.

UITS data centers are in the Computer Center with 1325 kW of battery backup and a 1750 kW generator for backup power. Cooling in the RDC is both in-rack cooling with chilled water heat exchangers and 70 tons of Computer Room Air Conditioning (CRAC) units. The UA Data Center is cooled with chilled water CRAC units. Both data centers are equipped with 18" raised floors that allow full coverage of cooling to all the equipment and equipped with leak detection systems in the subfloor. The fire suppression system is a multi-tiered defense with clean agent compressed gas, dry pipe pre-action sprinkler and EPO (Emergency Power Off) systems zoned to deploy in affected areas. UITS data centers have badge swipe access with two-factor authentication and video surveillance in the data centers and throughout the building. The data centers are monitored by a co-located 24/7 Operations staff and dedicated infrastructure team. Automated environmental and system monitoring assist with issue triaging and escalation. All personnel with access to unescorted data centers have undergone background checks and are required to be US Citizens.

1.4 Regulated Research Data and Computing Environments

Legal Aspects of Regulated and Controlled Data

Last Update:

Point of Contact: John Howard, HIPAA Compliance Office, jfhoward@email.arizona.edu

Regulated Research Environments (CUI, HIPAA, etc)

Last Updated: 2020-1-08

Point of Contact: Ryan Duitman (rduitman@email.arizona.edu)

The University of Arizona Information Technology Services department provides and manages secure research environments within the AWS GovCloud service, available to UArizona researchers to support security requirements for CUI and CUI / Export Controlled research projects. A small pilot will be launched Q1 2020 to support research using protected health information regulated by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) resulting in a late spring / early summer 2020 service launch of the environment for projects under the provisions of a Data Use Agreement (DUA) but are not required to meet the additional requirements that pertain to data used as part of a standard of care practice. These environments will provide researchers computational and storage resources for active regulated data analysis. Management of the environment and security controls as well as technical support will be centrally funded.

Services provided:

- Compliant computational and storage resources for active regulated data analysis
- Project planning and budgeting
 - Consulting regarding fit and needs of project
 - Proper sizing of computation and storage resources
 - Analysis regarding the computational and storage resource costs associated with project
- Deployment and decommissioning of computational and storage resources
- Environment account access and management
- Technical support for resources provided by the service

1.5 CyVerse Computational Resources

Last Updated: 1-08-2020

Point of Contact(s): Nirav Merchant(nirav@email.arizona.edu), Edwin Skidmore(edwin@cyverse.org)

Funded by the National Science Foundation's Directorate for Biological Sciences, CyVerse provides scientists with powerful computational infrastructure to handle huge datasets and complex analyses, thus enabling data-driven discovery. CyVerse's powerful extensible platforms provide data storage, bioinformatics tools, image analyses, cloud services, APIs, and more.

CyVerse computational grid clusters (1500 processing cores), virtualization clusters (Atmosphere cloud service platform) servicing more than 3000 virtual machines with 8TB RAM (maximum resources per instance is 32-cores and 256G RAM), CyVerse's data services, and 7 PB of storage are located in two data centers at UA: the BIO5 Institute and the UITS Data Center. Both facilities have redundant internet and power. These facilities are connected with high performance 10G networking. A third facility is used for disaster recovery and monitoring of CyVerse services located at the Texas Advanced Computing Center (TACC) in Austin, Texas. TACC stores a full backup copy of the entire CyVerse Data Store, updated nightly. The CyVerse Collaborative and TACC's facilities have 10G connectivity using Internet2 via Sun Corridor networking backbone, a necessity for the large-scale data transfers to supercomputers at TACC. TACC resources include:

- 18 petaflops of peak performance
- 4,200 Intel Knights Landing nodes, each with 68 cores, 96GB of DDR RAM, and 16GB of high speed MCDRAM
- 1,736 Intel Xeon Skylake nodes, each with 48 cores and 192GB of RAM
- 100 Gb/sec Intel Omni-Path network with a fat tree topology employing six core switches
- Two dedicated high performance Lustre file systems with a storage capacity of 31PB
- TACC's Stockyard-hosted Global Shared File System provides additional Lustre storage

If additional resources beyond UA and CyVerse are needed, the XSEDE national supercomputer network, which includes the Jetstream cloud-based computing platform may be employed.

1.6 XSEDE and Open Science Grid (OSG) National Computing Resources

Last Updated: 2020-01-08

Point of Contact(s): Blake Joyce (bjoyce3@email.arizona.edu)

If additional computing resources beyond UA and CyVerse are needed, the XSEDE national supercomputer network and the OSG are available to any United States-based researcher. These resources are freely available through an allocation request. See <https://www.xsede.org/web/site/for-users/getting-started> for more information.

The UA Research Computing department is committed to helping UA researchers apply for allocations, access the XSEDE and OSG resources they receive in allocations, and generally support researchers leveraging these resources. To that end, UA Research Computing is a member of the XSEDE Resource Allocation Committee (XRAC) which meets four times a year to review and award resource allocations to researchers.

CyVerse facilitates collaboration and access to more powerful computing resources provided by the national Extreme Science and Engineering Discovery Environment (XSEDE) network.

1.7 UITS Network Capacity

Last Updated:

Point of Contact(s): Dragana Vasic (dvasic@email.arizona.edu)

In addition to direct connections to commodity Internet service providers, UA is a member of Internet2 (I2). I2 is a not-for-profit organization that has built a 100Gb (gigabit) ultra-high speed backbone network for advancing Research and Education nationwide. The UA connection to I2 is through the Sun Corridor Network – a regional network established through a collaborative effort by Arizona State University (ASU), Northern Arizona University (NAU), and UA. The Sun Corridor Network provides advanced networking services beyond those available from the individual Arizona Universities and builds an environment essential to leading-edge education, research, and the sharing of digital communications resources, network services, and applications among eligible members. UITS manages and operates the Sun Corridor Network. The current connection from UA to Sun Corridor is a primary 100Gb circuit and a secondary 10Gb circuit via two(2) redundant edge routers, while Sun Corridor is connected to I2 via dual 100G connections in Tucson and Phoenix. The UA campus network consists of eight(8) redundant core routers in two core sites and eight(8) distribution routers in three(3) distribution locations. The core and distribution layer devices are interconnected at multi-100Gbps and most campus buildings are connected at 10Gbps. All buildings on campus are dually connected to campus distribution locations. Most customer ports are 1Gbps.

Researchers may also use a Science DMZ for fast high volume data transfers to outside collaborating institutions. The Science DMZ is deployed at the university network perimeter, outside border firewalls, via direct connections to a campus DMZ/Edge at multi-10Gbps. It is secured via static access lists deployed on two redundant, high throughput L3 switches dedicated to Science DMZ. There are two high-performance Data Transfer Nodes deployed in the Science DMZ managed by UITS.

The UA Research Data Center provides two(2) points of 40Gbps connections to the UA Data Center Fabric with all the servers connected by 1Gbps, 10Gbps, and 40Gbps connections. Servers are connected to two Cisco FEX switches for redundancy, as well as two Dell Z9500 high-density switches. Each pair of switching environments uplink to campus data center via four(4) 10Gbps circuits. This Data Center is located within the University network and is protected by both the campus border next-gen firewalls and by additional firewalls specific to enterprise and research Data Center services, providing higher-security for sensitive data than that available in high-speed Science DMZ.

The University of Arizona network is IPv6-capable with IPv6 routing enabled throughout the central network infrastructure. IPv6 service is enabled for departments and campus units upon request. The University of Arizona holds a /32 address allocation from ARIN and maintains native IPv6 connectivity to both Internet2 and the commodity Internet.

1.8 Data Science Institute (Data7)

Last Updated: 2019-11-15

Point of Contact(s): Nirav Merchant (nirav@email.arizona.edu), Maliaca Oxnam (maliaca@email.arizona.edu)

The UA Data Science Institute (Data7) was formed in 2017. Data7's goals include fostering the next generation of data-driven research by encouraging university-wide interdisciplinary collaboration, gaining external visibility, developing industry alliances, and increasing funding for UA research. By connecting UA researchers and aligning institutional expertise, computational resources, and infrastructure, Data7 enables investigators to ask more complex questions and achieve outcomes not easily attainable as individual investigators or within purely disciplinary teams. Data7 is also connected with the UA Transdisciplinary Research in Principles of Data Science ([TRIPODS](#)) effort, focused on the theoretical foundations of Data Science. Data7 actively coordinates and communicates training activities in cooperation with the [Data Science Resources & Training \(DSRT\)](#) group.

1.8.1 Project-based research support

Data7 solicits [one-page white papers](#) describing projects that fit the mission and goals of Data7 and fall into our [focus areas](#):

- **Machine Learning**
- **Natural Language Processing**
- **Image Analysis**
- **Large-Scale Data Visualization**

Projects should combine medium- to large-scale, multidisciplinary research teams with the intention of pursuing external funding in six to 12 months. The purpose should be to produce proof of concept and results that provide competitive advantage when pursuing funding opportunities.

Pilot projects include Natural Language Processing for data collected from telephone conversations between cancer survivors and health coaches, medical diagnostic image classification using Machine Learning, wildlife camera image analysis, and support of work to classify transient events observed in galaxies.

1.8.2 Data Science Training

Data Science is an interdisciplinary endeavor, merging techniques from computer science, statistics, and math with domain-level concepts to increase discovery through data analytics and visualization. However, taking advantage of advanced computational resources and the "big data revolution" requires a set of specialized skills and core competencies.

Data7 works closely with the broader campus community to support a variety of learning activities. Examples include Data and Software Carpentry training, vendor-provided training (Cloud, Machine Learning, GPU etc.) in collaboration with [UITS](#), and organizing special topic workshops (Containers, Workflow Managers). To foster Data Science Literacy, Data7 partners with the Data Science Research & Training ([DSRT](#)) group, developers of the Data Science Information Portal ([DSIP](#)), an online hub enumerating computational resources, expertise, and training events, and providing easy-to-use tools for UA researchers to find resources of interest.

Data7 is also connected with the UA Transdisciplinary Research in Principles of Data Science ([TRIPODS](#)) effort, focused on the theoretical foundations of Data Science.

Data7 postdocs and technical staff in collaboration with DSRT and interested UA departments organize seminars and symposiums on Data Science-related topics. Data7 maintains an “Events” calendar to advertise training events and a “Classifieds” page that allows colleagues to to offer or ask for help with Data Science related issues.

1.8.3 KMAP Knowledge Map

The Institutional Knowledge Map (KMAP <https://kmap.arizona.edu/>) is a software platform for providing succinct and interpretable views of enterprise-wide expertise and knowledge portfolios for any organization. The platform utilizes advanced machine learning (ML) techniques for synthesizing structured and unstructured data such as publications, proposals, biosketches, social media and web presence for individuals in an organization, to construct a multi-dimensional institutional collaboration and knowledge graph. The resultant knowledge graph is readily interpretable and lends itself for data exploration with its advanced web-based visualization techniques that transform the graph relationships into geographical constructs (maps), highlighting clusters of novel expertise and collaborative activities that are hard to discover. Using advanced NLP (Natural Language Processing) techniques, it can identify potential collaborators and subject matter experts given text input such as calls for proposals. KMAP can be used to build transdisciplinary teams, identify subject matter experts, and elucidate complex relationships within large, diverse, and geographically distributed organizations such as UArizona.

1.8.4 Course Catalog Search

Data7 has worked with the UA Analytics & Institutional Research team to provide a tool that can be used to filter and search the UA Course Catalog for Data Science related coursework. Courses can be filtered by level, and keyword search can be used to find course titles or departments.

1.8.5 Center of Excellence in Data for Society (CEDS)

Data7 supports the CODATA-UA Center of Excellence in Data for Society (CEDS): a non-profit, non-partisan think tank created to convene stakeholders and explore the social impact and policy role of data on society. CEDS advances knowledge and contributes to positive policy outcomes through a robust portfolio of projects housed within [Program Offices](#), which leverage the expertise of our global network and reflect the talent of the University of Arizona's research community. [CEDS Program Offices](#) organize studies and conferences to:

- Advance data diplomacy scholarship infrastructure
- Conduct policy research
- Convene stakeholders to explore issues and advance solutions
- Coordinate science diplomacy certification
- Offer guidance to policymakers, faculty, and students on big data policy issues
- Host events
- Produce knowledge
- Provide career training to students and professionals in data policy and data diplomacy
- Publish findings and white papers

1.9 TRIPODS Initiative

Last Updated: 2019-11-15

Point of Contact(s): David Glickenstein (glickenstein@math.arizona.edu), Stephen Kobourov (kobourov@cs.arizona.edu), Joe Watkins(jwatkins@math.arizona.edu), Helen Zhang (haozhang@email.arizona.edu)

TRIPODS is an initiative funded by the National Science Foundation (NSF), and described by NSF as follows: “Data science is an important interdisciplinary field with significant impacts on many aspects of the modern world, including government, industry, academia, and the general public. **Transdisciplinary Research In Principles Of Data Science (TRIPODS)** brings together the statistics, mathematics, and theoretical computer science communities to develop the theoretical foundations of data science through integrated research and training activities focused on core algorithmic, mathematical, and statistical principles.”

UA is one of 12 initially funded TRIPODS institutions. As the nature of data becomes ever more complex, new fundamental advancements will arise out of collaborations between computer science, mathematics, and statistics. These advancements will call researchers and educators in each of these three disciplines to learn from each other and work together, developing new mathematical, statistical, and algorithmic principles which will strengthen the foundations of theoretical data science while driving discovery and innovation for emerging applications. The vision of UA TRIPODS is to provide the infrastructure and environment so that data scientists at the foundations will become vital partners on the journey to the new frontiers of knowledge. At the same time, UA TRIPODS aims to help establish and encourage a diverse population to pursue careers in quantitative and computational data science by offering a broad spectrum of research, education and outreach programs tailored to high school students through post-doctoral fellows.

UA-TRIPODS activities include weekly seminars, multi-day summer conferences (<https://sites.google.com/math.arizona.edu/tripodsummerconference2019/videos>), and sponsorship of the Stanford-affiliated “Women in Data Science (WiDS)” regional conference at UA. UA-TRIPODS also aims to cultivate new partnerships with local industries engaged in data-centric enterprises in astronomy, environmental science, genomics, lunar and planetary sciences, medicine, transportation, and optical sciences.

Research Working Groups will pursue fundamental questions in a number of different areas in the theoretical foundations of data science. By ensuring each group includes researchers from each area of the foundational disciplines, while also pairing these groups with science and industry partners, UA-TRIPODS will ensure that these projects will benefit from a truly transdisciplinary collaboration. Progress in collaboration will be measured by specialized program evaluation metrics. Yearly workshops, seminars, and brainstorming/visioning activities will enable UA-TRIPODS to lay the foundation for the future of training and research in data science. In addition, this project will support the development of a new undergraduate degree in Statistics and Data Science at the University of Arizona, expanded online offerings, and facilitation of cooperation in graduate education (both among the foundational disciplines and with the domain sciences). Students will be provided with opportunities to experience first-hand data analysis and interdisciplinary research and education. UA-TRIPODS will also engage partnering colleges and universities across the Southwest in the institute activities and then utilize these connections to support novel and effective inter-university and intra-university research and educational programs. Partnerships with Hispanic Serving Institutions and Tribal Colleges in southern California, Arizona, New Mexico, and western Texas will be strengthened.

1.9.1 TRIPODS+X

TRIPODS+X are partnerships between Science and Engineering Fields and the NSF TRIPODS Institutes. UA TRIPODS+X projects include Data Visualization “Data Science Pathways for a Vibrant TRIPODS Commons at Scale” and Research “ CollaborativeResearch: Multi-Level Graph Representation for Exploring Big Data”. TRIPODS+X has hosted a two-part “Lemon Labs” workshop in conjunction with the National Institute for Mathematical and Biological Synthesis (NIMBioS). The workshop was conducted in an innovation lab format, bringing together diverse participants to brainstorm and strategize broad solutions to varying questions. The 30 participants included researchers from data-driven open science disciplines, such as astronomy, earth sciences, computational and information sciences, mathematics, climate science, and cyberinfrastructure, as well as researchers with training and human-centered design expertise.

1.10 Research, Innovation & Impact

Last Updated: 2019-11-15

Point of Contact(s):

The **Office for Research, Innovation and Impact** (RII) supports the world-class research enterprise at the University of Arizona, which is ranked among the top 25 public universities nationwide with more than \$687 million in research activity. Our researchers continue to forge innovative pathways, form powerful collaborations and make remarkable discoveries.

The UA research community is committed to preparing students to thrive in a time of augmented intelligence and the fusion of the digital, physical and biological worlds. Our research is creating the next generation of knowledge—and is translated and applied to society’s grand challenges—to enhance economic development and promote a rich quality of life in Arizona and beyond.

Fundamentally, Research, Innovation and Impact is responsible for advancing transformative excellence in research across campus, with particular attention to our land-grant mission of service to the State of Arizona. We do so by enabling the research success of our faculty through supporting university research centers, institutes, museums and core facilities; providing research development, stewardship, compliance and safety services; and securing strategic external partnerships.

RII can help researchers with proposal development, finding funding, managing projects, and ensuring compliance with policies and procedures that promote the safe, legal, and ethical conduct of research. RII supports many Centers and Institutes and Core Facilities to support research. RII also provides the **Research Resources** gateway to information on [research services](#), [training & professional development](#), and [core facilities](#), as well as links to [forms & templates](#), [policy & guidance](#) and other useful information.

1.11 The Office of Digital Innovation & Stewardship (ODIS)

Last Updated: 2019-11-15

Point of Contact(s): Megan Finn Senseney (msenseney@email.arizona.edu)

The University of Arizona Libraries' [Office of Digital Innovation and Stewardship](#) supports the campus community in navigating a rapidly evolving digital landscape. We offer training and consultation on topics related to data-intensive research and digital scholarship, and we support open access and data sharing to advance global equity and the creation of new knowledge. Through partnerships and collaborations we preserve and provide access to distinctive digital collections and the scholarly record of the University of Arizona.

1.11.1 Computational Research Analysis and Methods (CRAM)

1.11.1.1 Computational Research Support Services

The Libraries currently hosts weekly drop-in hours for R and Python programming, GIS, and research data management where researchers can get one-on-one help in a supportive environment. We also provide topical workshops and events throughout the semester, and programming is listed at <https://new.library.arizona.edu/events>.

1.11.1.2 Research Data Management

Managing large quantities of research data is a challenge and requires a realistic assessment of the retention requirements with an understanding of available technologies and investments needed to support that retention. Archiving and sharing data depends on the data and the retention requirements for digital data preservation. Research data management support exists in various pockets across campus and is normally attached to a specific service or product. Campus-wide and discipline-agnostic data management support is currently centralized at the University of Arizona Libraries' Research Data Management Services.¹ The unit's mission statement:

"In support of research activities at the University of Arizona, the Data Management Services team aims to be a central point of service to meet student, staff, and faculty research data management and geospatial data needs. Through our no-cost consulting services, we support compliance with funder data management plan requirements, meeting data sharing mandates, increasing research efficiency and reproducibility, assisting faculty and students with GIS software, finding and accessing data (geospatial and otherwise), and more. We also provide referrals to campus partners and resources when appropriate. Finally, via our workshops and customized training services, we aim to foster Increased knowledge and practice of good data management principles."

The unit provides consultative services around

- Writing data management plans (DMPs) for grant applications
- Data management strategies and workflows for research projects and research teams
- Data Archiving, documentation, metadata, and publishing
- Supporting best practices in spreadsheets and databases
- Reproducible research

In addition, the unit directly supports:

- DMPTool (<http://dmptool.org/>): A tool to write data management plans with templates for many funders, including NSF.

¹ <http://data.library.arizona.edu>

- Open Science Framework (<https://osf.arizona.edu/>): A project and lab organization platform that integrates with many tools used in research (Google Drive, GitHub, etc) which facilitates collaboration and optional data sharing.

1.11.1.3 Research Data Repository

In an effort to foster open, reproducible, and collaborative research on campus through long-term stewardship, the Libraries are developing data repository services intended to provide responsible curation for research datasets and related supplementary research outputs (e.g., software/code, media) generated through research activities by University of Arizona-affiliated researchers. Initial implementation is focused on establishing an institutional instance of the Figshare platform and developing associated data-related services for the storage and public sharing of research data, code, and related research products produced by University of Arizona researchers and affiliates. It also includes the associated data management and other services needed for ensuring quality and ease of use (e.g., data curation, and integrations with other campus systems). Our interest is sharing research products that can be made public, and the data repository is not meant as a general data store for private information, data related to the business functions of the University, or any kind of Regulated data. We are currently seeking early adopters with data ready for deposit; interested researchers should contact Fernando Rios (fríos@email.arizona.edu).

1.11.2 Scholarly Communication and Research Information Management (SCRIM)

1.11.2.1 Scholarly Communication Services

The Libraries provides workshops and consultations on a range of issues in scholarly communication, including identity management, open access, fair use, and copyright. We offer assistance with ORCID and DOIs, and we also host a selection of scholarly journals edited and managed by UA students, faculty, and affiliates. For individual consultations, please contact Ellen Dubinsky (edubinsky@email.arizona.edu). Upcoming programming is listed at <https://new.library.arizona.edu/events>.

1.11.2.2 Campus Repository

On April 4, 2016, the University of Arizona Faculty Senate passed an [open access policy](#) that calls on the faculty and university to distribute faculty-authored scholarly articles to the widest possible audience through the [UA Campus Repository](#). Our institutional repository facilitates open access to the research, creative works, publications and teaching materials of the University through open licensing. We collect, share, and archive content that has been selected and deposited by faculty, researchers, staff and affiliated contributors, and deposited content is discoverable via Google, Google Scholar, and other search engines. Eligible content includes pre-prints, departmental publications, technical reports, white papers, theses, and dissertations. Contact repository@u.library.arizona.edu with questions.

1.12 The CATalyst Studios

Last Updated: 11-25-2019

Point of Contact(s): Jennifer Nichols, (jtn@email.arizona.edu)

The CATalyst Studios is located in the Main Library as part of the new Student Success District. CATalyst Studios features a Data Studio, Virtual Reality Studio and Maker Studio. All facilities

are open for use by all members of campus and the community. Learning and use of spaces is facilitated with workshops, user certifications and drop-in support.

1.12.1 Data Studio

The Data Studio features a 6x3 Data Visualization wall of Samsung UH46N-E displays, with a pixel count of 11520 x 3240.

1.12.2 VR Studio

The Virtual Reality Studio features the latest tethered and stand alone headsets, including Oculus Rift, HTC Vive, and Oculus Go, as well as stations for 3D modeling and a green screen cyclorama.

1.12.3 Maker Studio

The Maker Studio provides access to prototyping equipment for learning communities of all types. Technology includes 2 Epilog FusionPro laser cutters, Ultimaker 3 3D printers, Shapeoko CNC router, Roland Camm-1 Vinyl Cutter, Juki industrial sewing machine, arduino, raspberry pi. User needs regularly inform what technology and supplies are provided.

1.13 Advanced Visualization at the University of Arizona

Advanced data visualization:

- Data visualization wall
- VR/AR space

1.14 Health Sciences

The University of Arizona Health Sciences is the statewide leader in biomedical research and health professions training. The UArizona Health Sciences includes the Colleges of Medicine (Tucson and Phoenix), Nursing, Pharmacy, and the Mel and Enid Zuckerman College of Public Health, with main campus locations in Tucson and the growing Phoenix Biomedical Campus in downtown Phoenix. From these vantage points, the Health Sciences reaches across the state of Arizona and the greater Southwest to provide cutting-edge health education, research and community outreach services. A major economic engine, the Health Sciences employs nearly 5,000 people, has approximately 900 faculty members and garners \$200 million in research grants and contracts annually.

The University of Arizona Health Sciences is a major contributor to the UA's well-earned reputation as one of our nation's top research institutions. From investigating biological concerns at the molecular level to studying and understanding individual communities, UAHS researchers are achieving breakthroughs that will improve individuals' lives. Numerous [centers and institutes](#) for specialized research and patient care function under the auspices of the UA Health Sciences.

[The University of Arizona Center for Biomedical Informatics and Biostatistics \(CB2\)](#) provides a wide variety of [services and tools](#), as well as regularly scheduled [educational sessions](#). The StatLab provides [statistical](#) and [bioinformatics](#) consulting.

1.14.1 Surveys and Clinical Databases

[The University of Arizona Center for Biomedical Informatics and Biostatistics \(CB2\)](#) is pleased to be a partner in the Research Electronic Data Capture (REDCap) Consortium. The REDCap consortium is composed of thousands of active institutional partners in over one hundred countries who utilize and support REDCap in various ways. It supports a secure web application (REDCap) designed exclusively to support data capture for research studies and institutional projects. The REDCap application allows users to build and manage online surveys and databases quickly and securely. UA Health Sciences (UAHS) currently hosts hundreds of production and development REDCap projects. [Periodic training on REDCAP](#) is available.

1.15 BIO5

The [BIO5 Institute](#) connects and mobilizes hundreds of world-class plant, animal, and human bioscientists, engineers, physicians, and computational researchers to develop creative solutions for complex challenges such as disease, hunger, water and food safety, and other health issues facing Arizona.

This interdisciplinary approach has resulted in improved food crops, innovative diagnostics and devices, disease prevention strategies, and promising new therapies.

BIO5 initiatives and projects are carefully chosen to align with areas of state and national need and for which University of Arizona faculty already have significant expertise. This strategy catalyzes the capacity to expand impact, economic opportunity and external funding opportunities. BIO5 aims to harness the collaborative power of our five core disciplines – Agriculture, Engineering, Medicine, Pharmacy, and Science - to find bold solutions to complex, biology-based challenges affecting humanity. The BIO5 Institute has a long-held tradition and culture of [honoring the land-grant mission](#) of the University of Arizona, in addition to an emphasis on science outreach, community engagement, positively impacting our state, and inspiring the next generation of scientists.

BIO5 provides [funding support](#) and a wide variety of [support services](#) to UA researchers.

2. Improving Institutional Efficiency

Last Updated:

Point of Contact(s):

For the past 2 years, UA has been increasingly leveraging cloud services to reduce the need to maintain on premises servers. Key examples are: Student email services have been migrated to Google, Box online storage and collaboration tool has been licensed, UAccess Research Gateway launched in Amazon Web Services (AWS) in April 2015.

UA is home to many national and international research efforts that have developed significant CI dedicated for their projects and research communities, in collaboration with a global community of researchers from CISE (Computer Information Sciences and Engineering) disciplines these projects are continuously advancing CI capabilities. Presence of these projects, their active CISE collaboration and engagement with the campus infrastructure groups has helped guide the UA IT Strategic plan, incorporating some of the best practices and technologies from these projects to widely benefit UA researchers. This synergistic activity has ensured efficiency and informed decision-making as many of these projects rely on UA central IT capabilities and are able to scale their capabilities due to this active collaboration. Examples of upcoming activities include:

- Evaluation of a campus wide data management infrastructure using iRODS
- Utilizing Cloud computing based infrastructure to establish a statewide geospatial data sharing grid (AEGIS: Arizona Environmental Grid Infrastructure Service)

3. Developing Software Tools

Last Updated:

Point of Contact(s):

A crucial aspect of building a consolidated campus cyberinfrastructure is adopting and developing a common set of tools, applications, and middleware. These should be compatible with national cyberinfrastructure platforms ensuring robust and secure identity management with appropriate access and authorization capabilities, and integration with regional and national initiatives. To this effect UA has implemented its federated identity management (IdM) infrastructure using Shibboleth (Internet2) and supports enterprise wide single sign-on using Central Authentication Service (CAS); all campus Resources requiring IdM make use of this institutional infrastructure. Campus researchers are provided with technical support, sample source code and guidance to integrate their application with the institutional IdM. UA is an active member and contributor to both Internet2 and JASIG initiatives and UITS implements many software components developed by these organizations. We are part of national efforts such as Federated InCommon and CILogon. Relying on these enabling CI components, UA researchers are able to securely and reliably connect with multiple internal and external resources.

4. CI Support and Training

Last Updated:

Point of Contact(s):

An important focus for the CI plan is continuing education, training and outreach activities aligned to its CI vision. Currently these include for credit course work (graduate and undergraduate) offered through multiple departments and special topic workshops. In order to ensure successful and effective utilization of CI, researchers are provided access to technical support infrastructure that includes wikis, email lists and in-person consultation with technical staff. With greater access to distance education such as workshops provided through NSF XSEDE and vendor provided training, UA CI support and training effort will further incorporate these opportunities with in-house capabilities, with the goal of CI aligned workforce development. UA campus hosts regular workshops in collaboration with Software Carpentry and Data Carpentry foundations and projects such as CyVerse to educate researchers on best practices and establish a well informed community of practice on campus.

4.1 Data Science Resources & Training

Last Updated: 2019-11-15

Point of Contact(s): Jeff Oliver (jcoliver@email.arizona.edu), Vignesh Subbian (vsubbian@email.arizona.edu)

The rapidly expanding data science endeavor is quantitatively and qualitatively transforming research. Improved compute and storage, combined with novel approaches and algorithms provide researchers a number of tools for predictive analytics, classification, and visualization. This revolution is largely benefiting from the application of advances in statistics and computer

sciences to specific domains in the social and natural sciences. Such approaches allow researchers to extract knowledge from large data sets and address previously unasked questions. However, a significant majority of students and faculty lack sufficient data science literacy skills to take advantage of the expanding landscape of data science tools.

Support for development in these skills and models is generally lacking in most formal training programs outside of computer sciences, despite a demand across the sciences for this training, especially in research computing. In many cases, these training demands are outstripping the currently accessible resources, creating a critical gap between the needs of scientists and their respective curricula. Not only does this knowledge and skills gap prevent researchers from applying data science approaches to their work, it limits the types of questions they even know to ask. Many researchers are restricted by this gap, yet there are more or less isolated pockets of data science expertise within each college. Researchers needing data science training would benefit from collaborations and connections with those data science practitioners in their respective domains.

The University of Arizona provides many computational resources and training opportunities in skills necessary critical for data science. The Data Science Resources & Training (DSRT) group was established in 2017 to ensure these opportunities are exploited by the broad audience of researchers at the University of Arizona. The DSRT steering committee advocates, advises, and cultivates new data science opportunities at the University of Arizona. Members of the steering committee connect the University of Arizona research community with data science training opportunities and campus resources by:

- Actively promoting data science resources within their respective professional communities on campus
- Promote training events and information on campus resources
- Collecting feedback and reporting back on training successes and gaps in opportunities
- Identifying additional data science trainers and opportunities

4.1.1 Data Science Ambassadors

Last Updated: 2019-11-15

Point of Contact(s): Jeff Oliver (jcoliver@email.arizona.edu), Vignesh Subbian (vsubbian@email.arizona.edu)

One specific effort of DSRT, the Data Science Ambassadors (DSA) program, was developed to help meet the growing demands for data science training across the University of Arizona campus. The program aimed to create accessible support within individual colleges by capitalizing on domain knowledge and data science expertise in the graduate student population. Graduate students' ability to "speak the same language" as researchers in their respective college reduces the communication barrier many researchers face when learning data science applications. The ultimate goal of the program is for graduate students to assist researchers and educators in navigating data science campus resources and to provide discipline-specific data science training.

The 2018-19 cohort of eight ambassadors engaged in a number of data science training events. Much of the efforts were through (1) Carpentries workshops: four of the ambassadors served as helpers for three Carpentry workshops and (2) "Astro Code" workshops organized an Ambassador. The three Carpentries workshops provided training to a combined 143 members of the UA campus. The five Astro Code workshops had an average attendance of 13 participants, and featured workshops on parallel processing and container technology. Two of the ambassadors were instrumental in organizing the Women in Data Science Tucson event on

5 April 2019. Over 100 people attended the event, which showcased numerous data science efforts on the UA campus. The 2019-2020 cohort of 10 ambassadors represents a growth in the number of colleges with an ambassador (six colleges, up from three). All but one of the ambassadors completed the Carpentries Instructor Training program as part of their ambassadorship.

Supported in part through funding from the Data Science Institute, the primary goal of the DSA program is to improve data science literacy at the UA campus. Beneficiaries include the college the ambassador serves and the ambassadors themselves. Colleges benefit from having a single point person to assist in training and navigation of campus data science resources. Actively countering the “silo” effect, ambassadors will create bridges among campus units, affording their respective colleges greater connectivity to the data science community. Ambassadors will immediately benefit from the training curriculum, as badged Carpentries instructors are increasingly recognized as top-quality candidates in job searches. The network of ambassadors and data science practitioners will also provide opportunities for ambassadors to improve their skills and understanding of data science applications.

4.1.2 Data Science Fellows

Last Updated: 2019-11-22

Point of Contact(s): Nirav Merchant (nirav@email.arizona.edu), Eric Lyons (ericlyons@email.arizona.edu)

The University of Arizona Health Sciences (UAHS) Office of Data Science Services (ODSS) Data Science Fellows initiative is starting in early 2020. The DSF program will consist of cohorts of postdocs with a mix of expertise in data science techniques (machine learning, image informatics, and natural language processing). Additionally, the DSF program will average six people per cohort and the Project Coordinator will coordinate the day-to-day operations and serve as the main point of contact for the scholars, program development team, mentor team, and other stakeholders involved in the program.

4.2 Research Bazaar Arizona (ResBazArizona)

Last Updated: 2019-11-15

Point of Contact(s): Blake Joyce (bjoyce3@email.arizona.edu)

Research Bazaar is a worldwide festival promoting the digital literacy emerging at the center of modern research. Events occur year round on campuses around the globe. The UA ResBaz team is comprised of scientists, students, and researchers who can work with other UA affiliates to provide training and networking events that increase knowledge and boost productivity.

Regular UA events hosted by the ResBaz team include weekly “Coffee and Code” sessions, “Hacky Hour” meetings where anyone is welcome to come and learn about programming, databases, scaling code, in a friendly and creative environment. The ResBaz group also runs the Tucson Python Meetup where all python enthusiasts are welcome regardless of expertise level.

Each year a free, multi-day intensive conference (ResBaz Tucson) is organized to allow researchers to come together to up-skill in “next generation digital research tools and skills.” In the spirit of a marketplace or bazaar, ResBaz is a highly participatory event where researchers from many different disciplines can learn, share knowledge and skills, and have fun. This event grows in popularity and impact with each iteration.

5. Support for NSF Ten Big Ideas

Last Updated: 2019-11-25

Point of Contact(s): Nirav Merchant (nirav@email.arizona.edu)

5.1 [Future of Work at the Human-Technology Frontier](#)

Sensor technologies to monitor and enhance health in the workplace are becoming more important and more sophisticated. Data7 partnerships include groups in the UA Health Sciences, the Center for Integrative Medicine, and the Institute for the Environment. These groups collect and analyze data from individuals as well as office and home environments to study stress and promote well-being.

The UArizona College of Social Behavioral Sciences offers a graduate Computational Social Sciences Certificate to create an environment for social science Ph.D. students to acquire theoretical knowledge and mastery of computational skills necessary for analyzing large and complex data. This program is broad in scope with coursework offered by many departments including Sociology, the School of Information, and the School of Public Policy.

5.2 [Growing Convergence Research](#)

The UA Data Science Institute and CyVerse groups work closely with UA Health Sciences (UAHS.) The UAHS Strategic plan includes a “Health Analytics Powerhouse” to harness the power of large datasets to drive new research and improved health outcomes by developing centralized sources of health data (e.g., CyVerse, network science), improving secure ease of access to data (e.g., Banner clinical data), and building advanced analytics capabilities (e.g., Data7.)

The College of Agriculture and Life Sciences (CALS) supports Data Science Services and Training to enable researchers to use data to transform agriculture. The mission of the UA ag-data group is to provide scientists and engineers with open software, data, and computing that will allow more efficient discovery and invention so that we can engineer crops and manage sustainable agricultural landscapes that produce food, energy, and ecosystem services.

5.3 [Harnessing the Data Revolution](#)

Many schools, departments, and institutes are working to accelerate data-intensive research and improve education around Data Science. Examples include:

- Data7/DSRT
 - The goal of Data7 is to foster the next generation of data-driven research by encouraging university-wide transdisciplinary collaboration, gaining external visibility, developing industry alliances, and increasing funding for UA research. By connecting UA researchers and aligning institutional expertise, computational resources, and infrastructure, Data7 enables investigators to ask more complex questions and achieve outcomes not easily attainable as individual investigators or within purely disciplinary teams.
 - Examples of Data7 collaborations include:
 - Support of a Data Science Fellow who leads the Computations and Software Working Group for the [Event Horizon Telescope](#) project

- A pilot project with the School of Nursing and the School of Information to use Natural Language Processing to explore whether cancer survivors can be motivated to adopt healthy behaviors through individual coaching
 - Machine Learning classification of Medical Diagnostic Images and Agricultural Crop Images
- Data7/DSRT also support yearly cohorts of Data Science Ambassadors, graduate students selected to help researchers become more well-versed in Data Science. Periodic training events such as Software/Data Carpentry and Container Camps are also offered to the campus.
- CyVerse (<https://cyverse.org>)
 - CyVerse's key services for managing the entire data life cycle, from storage to analysis to publication to reuse and sharing, lets you manage all your data in one place. CyVerse helps researchers broaden data discoverability and enables users to seamlessly use larger-scale national compute resources.
 - CyVerse provides regular trainings around a variety of data-centric topics.
- UA Libraries provide Research Data Management Services and Tools such as:
 - Data Management Plan templates
 - Project Management and Collaboration Tools
 - Data Publishing and Retention guidelines and Best Practices
- University Information Technology Services (UITS)
 - UITS offers several [research computing services](#) co-funded by the [Office of Research, Innovation & Impact](#) (RII) and available to all members of the UA campus community. These include high performance computing and storage available at no charge in the Research Data Center and server housing space available in the Research Colocation Data Center. In addition, free software, tools, and consulting are available for research computing, data visualization, and statistics.
 - UITS sponsors training on the use of High Performance Computing (HPC) systems, as well as consulting on visualization and statistics.
 - UITS holds an annual IT Summit to allow all campus IT practitioners to network and learn about the latest topics in the field.
- Computer Science
 - UA [Computer Science researchers](#) are actively making advances in:
 - [Artificial Intelligence \(Natural Language Processing, Machine Learning, Vision\)](#)
 - [Parallel, Distributed, and High Performance Computing](#)
 - [Data Visualization and Graphics](#)
 - Computer Science faculty are an integral part of the Transdisciplinary Research in Principles of Data Science ([TRIPODS](#)) effort
- iSchool
 - The School of Information is the interdisciplinary place to gain data science training, regardless of your major or field of study. Faculty across many U.A. departments have data science expertise (e.g., Biosystems Engineering, Computer Science, Ecology and Evolutionary Biology, Electrical and Computer Engineering, Management Information Systems, Mathematics). But our faculty in the School of Information are prepared to talk across disciplines to provide the basic data science tools every student needs today. That is, the School of Information provides a **methodological hub for interdisciplinary data science at U.A.**, training students and doing research on such activities as data curation, database management, information ethics and policy, along with statistical and computational data analysis. Research Areas include:
 - [Applied Natural Language Processing](#)
 - [Biological Informatics](#)
 - [Data Science, Social Network Analysis, Computational Social Science](#)

- [Information Collections, Libraries, Databases, Archives, & Ethics](#)
- [Internet of Things](#)
- [Machine learning, Artificial Intelligence, & Algorithmic Thinking](#)
- [Social Science & Internet Studies](#)
- [Virtual Reality, Video Game Design, & Human-Computer Interaction](#)
- Management Information Systems (MIS) supports the following Centers and Labs
 - [Artificial Intelligence Laboratory](#)
 - [Center for the Management of Information](#)
 - [Information Assurance and Security Information Center](#)
- Computational Social Science Graduate Certificate Learning Outcomes include:
 - Students need mastery of a computer language which enables them to utilize the computational tools available.
 - Students need to develop and demonstrate proficiency in at least one method of computationally-intensive data collection, extraction, or analysis. This would include mastery of machine learning, computational linguistics, network analysis, visualization data management, web scraping, agent based modeling, and advanced statistical modeling among other techniques.
 - Students need to understand CSS as a broad epistemology or approach to establishing credible knowledge, its history, and the ethical issues associated with CSS. They also need to be able to explain to non-experts the computational approaches they have learned and used in their research.
- College of Engineering
 - Taking on today's pressing challenges – intractable and emerging diseases, cybersecurity, aging infrastructures, big data management, and natural resources conservation, for example – requires expertise from many disciplines. So we focus our research on far-reaching, interdisciplinary team projects and connect our capabilities to the issues that most need our attention.
 - UA College of Engineering faculty and graduate students, and many undergraduates, are collaborating with researchers across the University of Arizona and around the globe on projects that improve quality of life.
 - Our research portfolio includes 100 projects with an annual expenditure of \$32 million. Major sources of funding include NASA, the Department of Defense, Department of Energy, National Institutes of Health, and the National Science Foundation. Among funders from private industry are Raytheon, Honeywell, Boeing, Texas Instruments, Edmund Optics, Ventana Medical Systems and Tucson Electric Power.
 - Focus Areas:
 - **Sustainability and Infrastructure** – water treatment and reuse, bioremediation, biofuels and renewables, energy efficiency and storage, autonomous systems, traffic and transit systems, cost estimation, data management, infrastructure networks
 - **Biomedical Systems and Devices** – sensors, imaging, biomaterials, wearable devices, mobile diagnostics and testing
 - **Defense and Homeland Security** – explosives detection, active flow control for air and space flight, robotics, cybersecurity, wireless communication, space object behavioral sciences
 - **Advanced Manufacturing and Materials** – materials testing in extreme environments, nanotechnology and metamaterials design, computational modeling to predict properties and build stronger materials, mining logistics
- College of Science
 - Since 2006, UA [College of Science](#) has provided an annual [lecture series](#) on a wide range of science topics that impact our community. Topics have included

cosmology, neuroscience, transformative science, life science, evolution and climate change.

- UA Continuing and Professional Education offers Coding and Data Analytics [Boot Camps](#), Courses, and Certificates

5.4 [Mid-scale Research Infrastructure](#)

The Office for Research, Innovation & Impact (RII) supports Research Development Services (RDS), which partners with UA faculty, their departmental/unit business office, and Sponsored Projects & Contracting Services (SP&CS) on the stages of the proposal lifecycle from Opportunity Identification and Assessment to Revision and Resubmission. Post-award activities are supported by departmental/unit business offices, SP&CS, and Research Compliance Services.

5.5 [Navigating the New Arctic](#)

As global climate change accelerates with increasingly substantial impacts on communities worldwide, the need to understand and make reliable projections of future climate becomes ever more imperative. The [National Science Foundation](#)-funded Southern Ocean Carbon and Climate Observations and Modeling, or [SOCCOM](#), project is addressing this need by deploying 200 robotic floats in the Southern Ocean surrounding Antarctica to capture real-time biological, geological and chemical (often called "biogeochemical") data. With powerful, high-performance computational infrastructure and data management provided by UA [Research Computing](#) and [CyVerse](#), simulation data from [Geophysical Fluid Dynamics Laboratory](#) (GFDL) now reside on a platform where collaborators can securely share data, models and resources. With access to GFDL's climate simulations, SOCCOM scientists were able to present their recommendations at a G7 Summit meeting, from which a panel of experts recommended moving ahead with plans to deploy a global array of biogeochemically sensed ocean floats.

5.6 [NSF 2026](#)

UA has embraced the exploration of research questions that are large in scope and cross multiple disciplines.

The [Office of Research, Innovation & Impact](#) (RII) supports several Institutes and Centers that are involved in research to create a better future:

- [Institute of the Environment](#), working to identify and advance innovative solutions to environmental challenges in Arizona and around the world.
- [Water & Energy Sustainable Technologies \(WEST\) Center](#), a leading-edge venue focused on advanced water and wastewater treatment and monitoring, alternative energy and related technologies. WEST is co-located with Pima County's Agua Nueva Water Reclamation Facility and is adjacent to reclaimed water recharge basins and constructed wetlands – all of which compose a Water Reclamation Campus. WEST is unique in its ability to conduct the translational research necessary to advance the sustainable technologies of the future. With large portions of the United States currently affected by drought, WEST Center is poised to answer some of the nation's biggest questions about sustainable water and energy use.
- [Institute for Energy Solutions](#) (IES), with a mission to provide a hub for cross-college energy science, policy and technology research and development, which benefits the

public and enhances the land-grant mission of the University of Arizona. The IES is a platform for energy experts, who work closely with industry, NGOs, government, and communities to address emerging energy challenges and opportunities. The institute promotes work to more reliably integrate renewable energy into the grid through the Power Forecasting Group and energy storage research initiatives; develop novel materials for energy conversion, energy-efficient water use and intelligent building technologies; support regional test beds and facilities and study the societal and economic impacts of technological advancements. IES works to advance research, outreach and educational activities at the University of Arizona, focused on creating safe, resilient, and secure Energy Solutions. The Institute for Energy Solutions builds on the success of the University of Arizona Renewable Energy Network (UAREN) to create an organization that expands the collaboration with others to respond to the deeply interrelated sustainability issues associated with the energy- water- food nexus.

- [Transportation Research Institute \(TRI\)](#) is equipped with the interdisciplinary expertise necessary to address the challenges of an ever-evolving transportation ecosystem. As an internationally recognized leader in the field, TRI works with researchers across campus and around the country and with public agencies and private industry to capitalize on and expand established expertise in emerging transportation systems and critical, enabling technologies. With the goal of developing and deploying the next generation of transportation systems to benefit Arizona and the nation, TRI pursues the financial, legal, and governance support and resources for the development of complementary and integrated policy necessary to advance the transformation of transportation in a modern society.
- [Biosphere 2](#) is one of the world's most unique facilities dedicated to the research and understanding of global scientific issues. The Biosphere 2 facility serves as a laboratory for controlled scientific studies, an arena for scientific discovery and discussion, and a far-reaching provider of public education. Its mission is to serve as a center for research, outreach, teaching and life-long learning about Earth, its living systems, and its place in the universe; to catalyze interdisciplinary thinking and understanding about Earth and its future; to be an adaptive tool for Earth education and outreach to industry, government, and the public; and to distill issues related to Earth systems planning and management for use by policymakers, students and the public.

UA Computer Science professor Mihai Surdeanu and geoscientists Barbara Carrapa and Mihai Ducea submitted a proposal “[Globalization of Science Results with AI](#)” to the [National Science Foundation’s 2026 Idea Machine](#) competition in 2019. From more than 800 submissions, their proposal was among 33 selected to continue to the second phase of the competition. Another submission selected for this second round was “[Reinventing Scientific Talent](#)” by [CyVerse](#) education, outreach and training lead Jason Williams.

The [Biosystems Engineering Department](#) is developing innovative systems to sustainably feed an increasing population through cutting-edge technology, sustainable practices, and the drive to make a difference in the world. We build sustainable solutions that create resilience among society, the environment and the economy. We are currently working on many projects, including optimizing growth conditions in controlled environment agriculture, finding alternative fuel and feed sources, irrigating crops sustainably, analyzing big data sets generated from using sensors and controls and looking at metagenomics

Tackling bold research problems will also necessitate the inclusion of talent from beyond traditional fields and engaging young students as potential future researchers. To promote STEM learning in a diverse population, RII includes a university-wide center that supports efforts to recruit and retain a diverse population of learners into a robust STEM pipeline. The [UA STEM Learning Center](#) offers consultations on broadening participation and broader impacts; project design; education research and evaluation; inclusive practices; and collaborating with

partners within the university, across institutions, or with industry. Resources on inclusive practices in STEM, current data for project proposals, and opportunities to network and collaborate with like-minded groups or individuals are available.

A team of UA faculty members and graduate students, including ECE professor [Kelly Potter](#), have joined forces with undergraduates on the Navajo Nation to [design a solar-powered water filtration system](#) that can provide 50 gallons of safe, clean water to 30 Navajo families per day. The students are part of the first cohort of trainees to participate in a five-year, \$3 million National Science Foundation grant that enables an unprecedented collaboration between the UA and Diné College, which serves a predominantly Navajo student population. The goal is to teach the next generation of STEM professionals how to confront food, energy and water challenges among indigenous communities while letting traditional values guide their work.

The Office of the CIO and UITs staff collaborate with members of the campus community on IT in multiple ways that promote innovation. In particular, UITs sponsors and supports:

- The UA [Information Technology \(IT\) Summit](#), an annual event that brings together the entire campus IT community to communicate, collaborate, and educate each other on current projects and best practices. The [IT Summit](#) is a day-long event that focuses on the power of collaboration among UA IT professionals, enhances partnerships with campus and industry partners, and showcases technologies that will transform the university's ability to innovate, collaborate, and educate for years to come. Recent Keynote speakers include Jaime Casap, Global Education Evangelist at Google, and creativity expert Tania Katan, an award-winning author, public speaker, and playwright, who teaches people and companies to generate imaginative innovations and breakthroughs.
- The [IT Leadership Academy](#), in partnership with the Division of Human Resources, a development program for UA IT professionals who want to become stronger, more effective managers and leaders.
- The [IT Student Advisory Board](#), which provides a means of communication and feedback with the UA student population around their IT needs.
- [UA Women in Technology](#) (UA WIT) is an inclusive community of women tech professionals from the UA campus and local Tucson community. UA WIT seeks to help empower women in tech fields through the offering of professional development opportunities such as workshops, structured mentorship, and networking events.

The [BIO5 Institute](#) supports the [KEYS Research Internship Program](#), a unique summer opportunity for motivated Arizona high school students with a strong interest in bioscience, engineering, environmental health, or biostatistics to work side-by-side with top faculty in University of Arizona laboratories.

The UA's Albert B. Weaver Science-Engineering Library hosts a number of [technology-rich collaborative spaces](#), including group study rooms, the University's first [collaborative learning classroom](#), [a Seed Library](#), and a [makerspace](#) with 3D printing.

5.7 [NSF INCLUDES](#)

In addition to the [UA STEM Learning Center](#), there are several groups on campus that promote inclusion and diversity:

- The [Commission on the Status of Women](#) is a well-established advocacy arm of the University of Arizona that actively champions a just, equitable and inclusive campus. We shape policy, address current needs and promote the advancement of all gender identities.

- The [Women in Science and Engineering \(WISE\) Program](#) was established at the UA in 1976 as part of the Women's Studies Department and the Southwest Institute for Research on Women. WISE aims to increase interest and diversity in the fields of social and natural science, technology, engineering, mathematics, and medicine by offering a variety of outreach programs and student engagement opportunities.
- [UA Women in Technology](#) (UA WIT) is an inclusive community of women tech professionals from the UA campus and local Tucson community. UA WIT seeks to help empower women in tech fields through the offering of professional development opportunities such as workshops, structured mentorship, and networking events.
- Beginning in 2019, the UA [Data Science Institute](#), Transdisciplinary Research in Principles of Data Science (TRIPODS), and [Center for Biomedical Informatics and Biostatistics \(CB2\)](#) sponsored a regional “Women in Data Science (WiDS)” Conference, in conjunction with the event that originated at [Stanford University](#). The Women in Data Science (WiDS) initiative aims to inspire and educate data scientists worldwide, regardless of gender, and support women in the field. Additional departments will be joining as sponsors for future UA WiDS events.
- The [Women in STEM Student Council \(WiSSC\)](#) is a student-directed council that supports the recruitment, retention, and graduation of female-identifying students pursuing degrees in STEM at the University of Arizona (UA). WiSSC serves as a network for the many existing “Women in STEM” graduate/undergraduate student groups and STEM departments and develops university-wide, coordinated programs. WiSSC also works with university administration at several levels to set university-wide priorities, identify areas of need, and inform policy to foster a diverse and inclusive STEM environment at UA that truly represents the demographics of the state of Arizona. **This initiative is the largest institutionalized university effort directed at supporting the entry and success of women in STEM fields in the United States. This makes UA a leader in creating the kind of diverse and inclusive STEM communities that are necessary for grappling with the pressing challenges of the 21st century.** This initiative is funded by the UA Office of the Provost and housed in the Women in Science and Engineering Program and Southwest Institute for Research on Women.

5.8 [Quantum Leap](#)

Researchers in the UA [Department of Materials Science and Engineering](#) have demonstrated the possibility for acoustic waves to do quantum information processing without the time limitations and fragility. The team is working with [Tech Launch Arizona](#), the office of the UA that commercializes inventions stemming from research, to patent their device and is investigating commercial pathways to bring the innovation to the public.

5.9 [Understanding the Rules of Life](#)

The [Office of Research, Innovation & Impact](#) (RII) includes the [UA Genetics Core \(UAGC\)](#) Facility, to provide researchers access to state-of-the-art resources and services to help investigators, educators, students and the biotech community conduct and promote research in the field of genomics. UAGC provides a wide range of molecular biology services and support ranging from Complete Solutions to fee-per-use services.

The [BIO5 Institute](#) at the University of Arizona was launched in 2001 with financial support generated by the Technology and Research Initiative Fund (TRIF), a special investment in higher education made possible by the passage of Proposition 301 by Arizona voters in November 2000. This tax was intended to expand major efforts in biomedicine and biotechnology in the state. The BIO5 Institute connects and mobilizes hundreds of world-class plant, animal, and human bioscientists, engineers, physicians, and computational researchers to

develop creative solutions for complex challenges such as disease, hunger, water and food safety, and other health issues facing Arizona. This interdisciplinary approach has resulted in improved food crops, innovative diagnostics and devices, disease prevention strategies, and promising new therapies. BIO5 initiatives and projects are carefully chosen to align with areas of state and national need and for which University of Arizona faculty already have significant expertise. This strategy catalyzes the capacity to expand impact, economic opportunity and external funding opportunities.

The BIO5 Institute has a long-held tradition and culture of honoring the land-grant mission of the University of Arizona, in addition to an emphasis on science outreach, community engagement, positively impacting our state, and inspiring the next generation of scientists. BIO5 scientists work with the UA's [Tech Launch Arizona](#) to commercialize discoveries and facilitate connections and collaborations among life scientists, companies, industry, entrepreneurs, venture capitalists, and economic development organizations. BIO5-affiliated researchers have fostered forty-five spinoff companies in the last decade, resulting in new technologies, diagnostics, and treatments.

[Planet Microbe](#) is a three-year project of the Bonnie Hurwitz lab, funded by the National Science Foundation to bring together genomic and environmental data sets coming from ocean research cruises. Samples of water are taken using an instrument that measures salinity, temperature, depth, and other features to create a scan of ocean conditions across the water column. As the instrument descends into the deep dark ocean, water samples are collected at different depths for a variety of experiments including sequencing the DNA/RNA of microbes. The project's data infrastructure is built on top of NSF's [CyVerse](#), also led by the UA, to integrate and build information from diverse data stores in collaboration with the broader cyber community.

The UA [Center for Biomedical Informatics & Biostatistics \(CB2\)](#) has a mission to enable data-driven healthcare and research to improve patient outcomes. Their strategies include building a data-driven learning health system and advance precision medicine analytics, promoting health equity through education, and providing data science resources and tools.

The UA [BIO5 Statistics Lab](#) is available for expert consultation on experiment design, bioinformatic analyses, grant preparation, and manuscript submission. They also provide drop-in clinics, open to all UA investigators in need of statistical guidance.

5.10 [Windows on the Universe](#)

The University of Arizona is a world leader in space sciences related research. There is a wide array of research interests reflected in the work of the Astronomy department, from the detection of new planets to measuring light from more than half a universe away. The department has a significant focus on the design and construction of cutting-edge instrumentation both for ground and space based facilities. The Department of Astronomy's academic program and the research prominence of Steward Observatory form one of the finest centers for astronomical studies in the world. Students have access to excellent observing conditions, world-class telescopes and a vibrant astronomical community. Astronomers here are among the national and international leaders in observational and theoretical research, making significant breakthroughs in related technology development from new light detectors to giant telescope mirrors.

The Extremely LarGe Advanced TechnOlogy (El Gato) cluster is a high performance computer jointly funded by the [National Science Foundation](#) and the [University of Arizona](#). The El Gato project is a collaboration between members of the UA [Department of Astronomy](#), the [School of Information](#), and the [University Information Technology Services](#).

UA is a key participant in the **Event Horizon Telescope** (EHT), an international collaboration capturing images of black holes using a virtual Earth-sized telescope. This worldwide effort helps astronomers study objects predicted by Einstein's theory of General Relativity. UA Data Science Institute Fellow CK Chan leads the Computations and Software Working Group for the EHT project. He [leveraged CyVerse Cloud Computing resources](#) in the development of one of the main data processing pipelines, and become the primary contributor to the major EHT software packages.

The EHT Theoretical Models and Simulations Working Group stores its simulation library in the [CyVerse Data Store](#), a secure, cloud-based repository for large-scale data storage and sharing. Influenced by CyVerse's education on best practices in data science, the working group has also begun to use the [Open Science Grid](#) to analyze its simulations.

6. Acknowledgments

This document was initially compiled and edited by Susan J. Miller, UA Data Science Institute. Other contributors, listed in individual sections, include:

- Blake Joyce, UA Research Computing
- Ryan Duitman, UA Research Computing
- Dragana Vasic, UA Research Computing
- John Howard, UA HIPAA Compliance Office
- Nirav Merchant, CyVerse/UA Data Science Institute
- Edwin Skidmore, CyVerse
- Eric Lyons, CyVerse
- Maliaca Oxnam, UA Data Science Institute
- Helen Zhang, UA TRIPODS Initiative
- David Glickenstein, UA TRIPODS Initiative
- Stephen Kobourov, UA TRIPODS Initiative
- Joe Watkins, UA TRIPODS Initiative
- Megan Finn Senseney, UA Office of Digital Innovation & Stewardship
- Jennifer Nichols, UA Libraries
- Jeff Oliver, UA Libraries
- Vignesh Subbian, UA Biomedical Engineering