# RII Track-1: Data Analytics that are Robust and Trusted (DART): From Smart Curation to Socially Aware Decision Making

# Supplementary Document: Gap Analysis, Needs Assessment, and Updated Cyberinfrastructure Plan

January 31, 2021


| | |
|---|---|
| Award Number: | 1946391 |
| Jurisdiction: | Arkansas |
| Start Date: | July 1, 2020 |
| End Date: | June 30, 2025 (Estimated) |

# Contents

1. Introduction and Background....4
1.1. Jurisdiction-Specific Terms and Conditions....4
1.2. Cyberinfrastructure Advisory Team....4
2. Gap Analysis and Need Assessment....6
2.1. Introduction....6
2.2. Recommendations and Initial Steps....6
3. CI Metrics and Governance....7
3.1. Cyberinfrastructure Metrics....7
3.2. Governance....8
3.2.1 Arkansas Research Computing Collaborative (ARCC)....8
3.2.2 Cyberinfrastructure Working Group (CWG)....8
3.2.3 Arkansas Research Platform (ARP)....10
3.2.4 Protected and Sensitive Information Storage....11
3.2.5 Tiered CI Planning....12
4. CI Learning Needs....13
5. Budget and Budget Justification....14
5.1.1 CI Budget Summary....14
5.1.2 Proposed Budget Modifications....14
6. Appendix A: CC* CIRA Proposal Development....16
7. Appendix B: Modified Activity Table....18
8. Appendix C: Modified Cyberinfrastructure Logic Model....24
9. Appendix D: Comprehensive "State of the Network" Across the Resource Providers....28

## List of Acronyms

| | |
|---|---|
| ACDS | Arkansas Center for Data Science |
| ADHE | Arkansas Department of Higher Education |
| AEDC | Arkansas Economic Development Commission |
| AI | Artificial Intelligence |
| ARCC | Arkansas Research Computing Collaborative |
| ARGO | The Great Plains Augmented Regional Gateway to the Open Science Grid |
| ARE-ON | Arkansas Research and Education - Optical Network |
| ARP | Arkansas Research Platform |
| ASU | Arkansas State University |
| BD2K | Big Data to Knowledge |
| CCPA | California Consumer Protection Act |
| CDS&E | NSF program: Computational and Data-Enabled Science and Engineering |
| CI | Coordinated Cyberinfrastructure Research Theme; also Cyberinfrastructure |
| CITI | Collaborative Institutional Training Initiative |
| COVID | Corona Virus Disease |
| CSTA | Computer Science Teacher Association |
| CUI | Controlled, Unclassified Data |
| DART | Data Analytics that and Robust and Trusted: From Smart Curation to Socially Award Decision Making |
| DC | Data Curation and Life Cycle Research Theme |
| DG | Data governance |
| DTN | Data Transfer Node |
| EAB | Employee Advisory Board |
| ED | Education Research Theme |
| EPSCoR | NSF program: Established Program to Stimulate Competitive Research |
| GDPR | General Data Protection Regulation |
| GPN | Great Plains Network |
| GPR CyberTeam | Great Plains Regional CyberTeam (NSF Award Abstract #1925681 CC* Team: Great Plains Regional CyberTeam) |
| GRA | Graduate Research Assistant |
| HBCU | Historically Black Colleges and Universities |
| HDFS | Hadoop Distributed File System |
| HIPAA | Health Insurance Portability and Accountability Act of 1996 |
| HPC | High Performance Computing |
| IHE | Institute of Higher Education |
| IUCRC | NSF program: Industry-University Cooperative Research Centers |
| LP | Learning and Prediction Research Theme |
| LSAMP | Arkansas Louis Stokes Alliance for Minority Participation |
| LSTM | long short-term memory |
| MTPP | marked temporal point process |
| NASA | National Aeronautics and Space Administration |
| NASEM | National Academies of Sciences, Engineering, and Medicine |
| NRT | NSF program: NSF Research Traineeship Program |
| ML | Machine Learning |
| MT | Management Team |
| OURRstore | The Oklahoma University (OU) & Regional & Research Store |

| | |
|---|---|
| PDC | Positive Data Control |
| PII | Personal Identifying Information |
| POC | Proof of Concept |
| REN-ISAC | Research & Education Networks Information Sharing & Analysis Center |
| RF | Random Forest |
| SA | Social Awareness Research Theme |
| SAC | Science Advisory Committee; also known as the Arkansas EPSCoR Steering Committee |
| SAU | Southern Arkansas University |
| SBIR | Small Business Innovation Research |
| SLURM | Simple Linux Utility for Resource Management |
| SM | Social Media and Networks Research Theme |
| SSC | Science Steering Committee; also known as the Leadership Team |
| SSP | System Security Plan |
| STC | NSF program: Science and Technology Centers |
| STEM | Science, Technology, Engineering, and Mathematics |
| STTR | Small Business Technology Transfer |
| SWOT | Strengths, Weaknesses, Opportunities, and Threats |
| UAF | University of Arkansas, Fayetteville |
| UALR | University of Arkansas at Little Rock |
| UAMS | University of Arkansas for Medical Sciences |
| UAPB | University of Arkansas at Pine Bluff |
| UCA | University of Central Arkansas |
| UGRA | Undergraduate Research Assistant |
| WCOB | Walton College of Business, University of Arkansas, Fayetteville |
| WD | Workforce Development |
| XSEDE | NSF program: Extreme Science and Engineering Discovery Environment |

# 1. Introduction and Background

## 1.1. Jurisdiction-Specific Terms and Conditions

This report is supplementary submission to the DART Strategic Plan submitted 10/06/2020 by the Arkansas EPSCoR jurisdiction. The special terms and conditions request that a technical advisory committee of experts with experience in planning future CI architecture specifically for data science shall be convened to assist the AR leadership and science team in developing data science CI plans that fully support the current and future research and education needs for this project. The output of convening the technical advisors with the project leadership will be:

1. ***A gap analysis and needs assessment for the project's research and education objectives, including specifics on handling of HIPAA-protected data.***
2. ***Details of CI metrics and governance processes for adjusting the CI investments as project needs change over the five years of this project.***
3. ***A plan for budget expenditures that addresses the gaps and needs appropriately.***
4. ***A plan to address DART project data science CI learning needs.***

Section 1 of this document describes the committee's evaluation of current gaps in the jurisdiction's CI architecture, with particular focus on the network architecture changes required among ARE-ON, UAF, and UAMS. Finally, Section 2 provides an overview of current CI metrics and the governance process for adjusting CI investments as project needs change over the course of the project. Section 3 addresses learning needs required to accommodate new data science clients. Finally, Section 4 describes the budget changes necessary to implement these changes outlined in the previous sections. Appendix A includes details of a CC* CIRA proposal to enhance the activities of planned jurisdictional meetings of working groups and advisory boards.

We note that some components of the CI Plan, especially elements of organizational structure, restructuring of visualization components, and CI training needs, were addressed in the Strategic Plan. We reiterate key elements of that plan and include, in Appendices B and C, updates to relevant entries in the activity matrix and logic model, respectively.

## 1.2. Cyberinfrastructure Advisory Team

Because of the close ties and involvement that Arkansas Universities have to the GPN, the jurisdiction contacted James Deaton, Executive Director of GPN, for assistance. Deaton suggested, and the jurisdiction accepted, assistance under the auspices of the Great Plains Regional CyberTeam (NSF Award Abstract #1925681 CC* Team: Great Plains Regional CyberTeam) with the follow principal investigators from across the region:

- Grant Scott GrantScott@missouri.edu (Principal Investigator)
- Timothy Middelkoop (Former Principal Investigator)
- Daniel Andresen (Co-Principal Investigator)
- James Deaton (Co-Principal Investigator)
- Kevin Brandt (Co-Principal Investigator)
- Derek Weitzel (Co-Principal Investigator)

The GPR CyberTeam is a perfect fit with the aims of the Arkansas jurisdiction in that they recognize that "advances in science and technology fields are increasingly accomplished as part of multidisciplinary and multi-institutional collaborations that require complex cyberinfrastructure." As described in its abstract, the project formed a:

> *"regional CyberTeam led by the Great Plains Network to support and advance the computational and data-intensive research across the region through the development of specific cyberinfrastructure resources, workforce training, and the development of unique, mutual, and cross-institutional support methodologies and agreements. The project advances the adoption and experience of advanced computing and data resources by developing a model built upon best and emerging practices for cross training and researcher outreach, pairing an experienced mentor at one institution with a mentee at another.*
>
> *The GPR CyberTeam project objectives are to:*
>
> - *Improve campus awareness and adoption of advanced cyberinfrastructure.*
> - *Increase the number of campus research computing and data professionals at mentored institutions, especially for institutions with small IT staffs with many job duties.*
> - *Increase the capabilities of campus cyberinfrastructure resources.*
> - *Enable development, deployment, and operation of cyberinfrastructure to make science efficient, trusted, and reproducible.*
>
> *The CyberTeam is a cross-institutional team consisting of technical leaders in the region paired with new members of the workforce, graduate and undergraduate students interested in joining the cyberinfrastructure workforce, and the institutional research computing leadership for regional research universities. It provides a model for distributed support teams to support cyberinfrastructure and aid in the development of a cyberinfrastructure engineering and facilitation workforce. Generalized best practices for a regional team of CI mentors including specific mentorship plans, retrospectives, and reference materials are disseminated."*

GPR CyberTeam CoPIs James Deaton and Kevin Brandt were directly involved in a series of meetings that included network engineers from ARE-ON, UAMS and UAF, security officers from UAF and UAMS, the CIO's and research associate CIO's from UAF and UAMS and the Executive Director of ARE-ON, and the three DART Research Theme Co-Leads. Details of these meetings are described in Section 2.

# 2. Gap Analysis and Need Assessment

## 2.1. Introduction

An initial gap analysis of the state of the CI associated with accomplishing the CI objectives of the DART project led to the identification of several areas in need of improvement. The current state for each of the institutions involved as resource providers in ARCC (UAF, UAMS, and UALR) were the focus with aspects of common needs providing early guidance for the other institutions in the project.

A comprehensive "state of the network" across the resource providers is already in early stages of development (Appendix D). Further collaboration and especially coordination among the ARCC resource providers and ARE-ON is needed to leverage common infrastructure to increase bandwidth and appropriate security of the paths while reducing the latency needed for aspects of the research testbeds. A research-based private network between institutions is a common regional approach for providing a more granular method of improving performance and simplifying access and data movement for researchers.

## 2.2. Recommendations and Initial Steps

**Recommendation 1:** Monitoring and measuring the capabilities of the current state of the network and as adjustments and upgrades are introduced needs to be implemented. Each of the individual universities and ARE-ON have practices in place to collect network telemetry but aspects need to be coordinated to provide a thorough view of the state of the network for ARP-related activities. In addition to the telemetry, additional perfSONAR nodes need to be deployed to assess the performance of the network and to gauge the impact of network changes. Such deployments need to be a coordinated activity to assure consistency in the testing processes and assure efficient operation and stable measurement archives.

**Recommendation 2:** The importance of federated identity practices grows as resources via ARCC are utilized across institutional boundaries. As resources are consumed (and shared) via the broader goals of ARP across state borders and nationally with the GPN Research Platform, Pacific Research Platform and XSEDE, InCommon membership and practices become very important. The state of federated identity at the institutions is mixed with only UAMS operating as an InCommon identity and service provider and none of the institutions registered as Research and Scholarship adopters. Efforts to address this should be guided by InCommon's Baseline Expectations for Trust in Federation Version 2 and REFEDS Research and Scholarship practices.

**Recommendation 3:** Review of the ARCC resource providers data controls included requests for documentation regarding NIST 800-171-related System Security Plans as well as regulatory compliance efforts associated with HIPAA and FERPA. Responses were mixed with nominal effort underway addressing university efforts toward dealing with CUI. The breadth of research to be addressed within DART, the diversity of participating institutions and the broader impacts of addressing a strategy for regulatory compliance make this project an intriguing potential engagement opportunity for Trusted CI. An upcoming engagement application window should be leveraged to garner the insight of this NSF Cybersecurity Center of Excellence to identify opportunities to address security and compliance aspects of the project. In addition, all the ARCC resource providers, ARE-ON and several of the other institutions are REN-ISAC members. REN-ISAC provides a peer assessment service which has recently shifted from an in-person endeavor to working remotely. It can also provide substantial insight in this area.

**Recommendation 4:** As resources are shared across institutions, local facilitation will play a valuable role. The XSEDE Campus Champions program provides a structure for the identification and a community of support for individuals serving in these roles. Aside from UAF, there are only 2 other

Champions identified within these institutions participating in the project, one at ASU and one at UALR. Individuals who will interface with researchers more directly need to be identified. The outreach and support of mentors within the GPR CyberTeam will work with these individuals to help identify more granular gaps in the CI as the project progresses.

**Recommendation 5:** Of the universities involved in DART, only UAF and UAPB have received funding from the NSF CC* program. All the other institutions remain eligible for funding within the program's area 1 and/or area 2. Significant improvements in research CI should be funded through this program and can occur in parallel with other activities within DART.

These recommendations are addressed in the sections following.

# 3. CI Metrics and Governance

## 3.1. Cyberinfrastructure Metrics

The current CI at UAF and UAMS was designed to service the dominant use patterns of large batch jobs, queued to run in SLURM, with user interaction provided by SSH. CI at UAMS is slightly more diverse but is none the less designed around gene sequencing and medical image processing. While the resulting computing patterns differed from UAF, access was still primarily through SSH and a scheduling service. Both architectures were facing pressure from data science-oriented users with their emphasis on machine learning from big data sets requiring more intensive and visual interaction with the compute nodes. A coordinated, SLURM scheduled, Open OnDemand service at both UAF and UAMS is already facilitating this higher level of interaction with its emphasis on Jupyter notebooks, R-Studio, and virtual machines.

*Table 1. Current CI metrics at UAF and UAMS.*

| Institution | Active users | CPU hours / year | Software patterns |
|---|---|---|---|
| UAF | 422 | ~50M | VSAP, Molpro, MDMPI, MKL with Python and far less R use |
| UAMS | 183 | ~ 60M | Bioconductor and Qiime 2, other bioinformatics applications in R and Python, |

The proposed research will be supported by a data science cyberinfrastructure (CI) platform capable of providing secure, distributed, agile, scalable, and on-demand services. We propose to architect and build a private cloud environment, the ARP (Figure 2) and integrate it with existing high-performance computing and petabyte scale storage resources. In combination, these will provide 1) libraries of pre-configured containers designed to support a variety of well-known and novel workflows in machine and statistical learning, graph theory, bioinformatics, and geoinformatics, 2) containers configured for parallel computation and distributed memory on HPC resources for analysis of very large datasets, 3) the ability for researchers to create and share new containers and share, and 4) the ability to move data to visualization environments to UALR visualization resources to aid in analysis and meta-analysis of experiments.

Section 3 addresses the learning needs associated with this new data science orientated architecture.

## 3.2. Governance

### 3.2.1 *Arkansas Research Computing Collaborative (ARCC)*

In May 2020, UAF and the UAMS entered a formal partnership to consolidate the management of high-performance computing centers at each institution into a single entity, ARCC. ARCC will be implemented in the first year of the grant and expanded to include resources available at the UALR Emerging Analytics Center. These three institutions will act as resource providers as well as consumers while the other institutions in DART will consume these resources with direct access to big data through Globus and associated data transfer nodes, code sharing through a private GitLab installation, and computing in interactive and batch sessions on existing and new computational nodes.

As described in the DART Strategic Plan, the CI for DART, dubbed the Arkansas Research Platform (ARP) will be managed as a unique multi-institutional resource by ARCC. ARP resources consist of computing resources managed by the high-performance computing centers at UA and UAMS and visualization resources at UALR. The DART Co-Leads will serve as the Leadership Team for this resource and direct the management of resources at their respective sites. Dr. Cothren will serve as the executive director of ARCC. The leadership team will define the operational procedures for the ARP combined resource in consultation with a user committee comprised of major users from each campus. A memorandum of understanding among the campuses participating in ARP will define the governance structure and establish operational parameters. This governance and operations model is based on our experience operating the established facilities at UAF, UAMS, and UALR. Direct support to research faculty and students will be provided by existing staff at UAF and UAMS, with additional faculty and students from UALR assisting in the development of testbed solutions.

### 3.2.2 *Cyberinfrastructure Working Group (CWG)*

A newly formed CI working group, chaired by James Deaton, executive director of GPN, organized and managed by UAF Information Technology Services, and composed of the GPR CyberTeam CoPIs, has been formed and will continue to advise ARCC in filling the gaps identified in the initial analysis and in ongoing analysis.

Table 2 identifies the individuals currently assigned to the CWG. The administrators, engineers and researchers are all involved in the design and implementation of the network architecture required to make the ARP available of more members of the jurisdiction. The CWG and inclusive subgroups have been meeting monthly since October 2020 to discuss current network configurations, rational for requested changes to that network, and the security impacts of the changes. One of the subgroups has specifically addressed plans to manage CUI data at UAF in accordance with 800.171.

The CWG will assist the ARCC and DART to eventually address all five recommendations. However, the makeup of this particular group is targeted to immediately address **recommendations 1, 2, and 3**.

*Table 2: The CWG is comprised of researcher and IT engineers from UAMS, UAF, ARE-ON. Membership on this team will change as new campuses develop CI plans consistent with ARCC plans.*

| Name | Role | Affiliation |
|---|---|---|
| James Deaton | Executive Director, CyberTeam Co-PI | Great Plains Network |
| Kevin Brandt | Director of Research Computing, CyberTeam Co-PI | South Dakota State University, Brookings, SD |
| Brian Berry | Administrator, MT | University of Arkansas, Little Rock |
| Jan Springer | Director of Emerging Analytics Center, DART CI Theme Co-Lead | |
| David Merrifield | Interim Executive Director | Arkansas Research and Education - Optical Network |
| Scott Gregory Ramoly | Chief Technology Officer | |
| Guy L Hoover | Manager of Network Engineering | University of Arkansas for Medical Sciences |
| Shawn Bynum | Director of Unified Communications | |
| Stephen Cochran | Chief Information Security Officer | |
| Matthew Reiss | Network Capacity Engineer | |
| Eric Wall | Assistant Director of IT Security | |
| Guy L Hoover | Manager of Network Engineering | |
| Fred Prior | Chair of Bioinformatics, DART CI Theme Co-Lead | |
| Stephen L. Tycer | Chief Information Security, | University of Arkansas, Fayetteville |
| Elon T. Turner | Network Director | |
| Lisa Richardson | Director, Project Management Office | |
| Michael E. Davis | Network Architect | |
| Nick Salonen | Senior Information Security Analyst | |
| James McCarthy | Project/Program Manager (Enterprise Services) | |
| Don DuRousseau | Associate CIO for Research | |
| Jackson Cothren | Director AHPCC, DART CI Theme Co-Lead | |

The overall DART management plan reflects the previous EPSCoR project organizational structure. It includes a state-supported Central Office which provides general oversight for the project and coordinates interactions with state boards and agencies; a management team (MT) comprised of administrators from participating campuses, Table 2, to ensure project implementation on campuses and information flow; a researcher-led Science Steering Committee (SSC) provides oversight for the scientific aspects of the program; and one or more external advisory boards contribute stakeholder perspective and facilitated dissemination of results to other groups. The CWG will report to the Scientific Steering Committee (SSC) directly and through the CI Research. Theme (Figure 1).

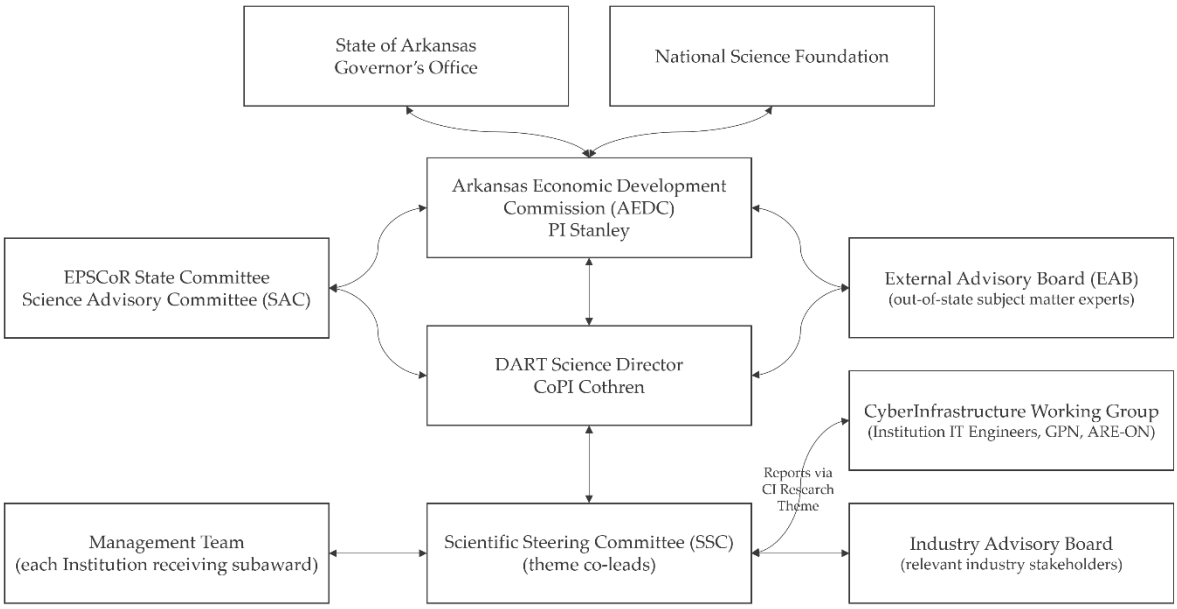


*Figure 1: Management structure of DART with the addition of the CI Working Group*

### 3.2.3 ***Arkansas Research Platform (ARP)***

The configuration and implementation plan of the ARP remain largely unchanged from the Strategic Plan with a few notable exceptions detailed here.

ARP will use Globus Research Data Management services for big data sharing and management across the seven research campuses. Endpoints will be set up at UAF, UALR, UAMS, ASU, UCA, SAU, and UAPB. The original plan called for the licensed Enterprise version of Globus for the full five years of the project. However, based on initial tests using the free services provided by Globus the CWG determined that the additional features were unnecessary in the first three years of the project at least and the funds set aside for the licenses were repurposed as described in Section 3.

Code archiving and sharing will use the existing GitLab for Enterprise installation at UAF.

ARP will support a variety of research computing platforms: traditional bare-metal HPC jobs, Singularity containers, and kernel virtual machines (KVMs) on the existing Pinnacle and Grace clusters as well as the new data science cluster funded as part of this proposal. The Sample Linux Utility for Resource Management (SLURM) scheduler will be used to provision all three types of jobs. Singularity containers are a widely accepted, secure standard in a multi-user HPC environment where access to the data of other users must be restricted. The container jobs will require a user to either download a container from external repositories like NVidia NGC (in native Singularity format) or Docker Hub (easily convertible to Singularity format) or use containers stored on shared local storage on Pinnacle or

Grace. Once the container is in place, a single-line command in the SLURM job script will bind the user's input data directory to the container and run the executable inside the container on the input data. The job terminates when either the executable in the container finishes processing the input, or the job exceeds the requested wall time. A variety of big data management resources such as HDFS and Apache Spark will be enabled using the MapR Sandbox and internet-facing graphical interfaces will be provided to run services such as Jupyter Notebook, RStudio Server, and the HPC scheduler, Open OnDemand, at both UAF and UAMS.

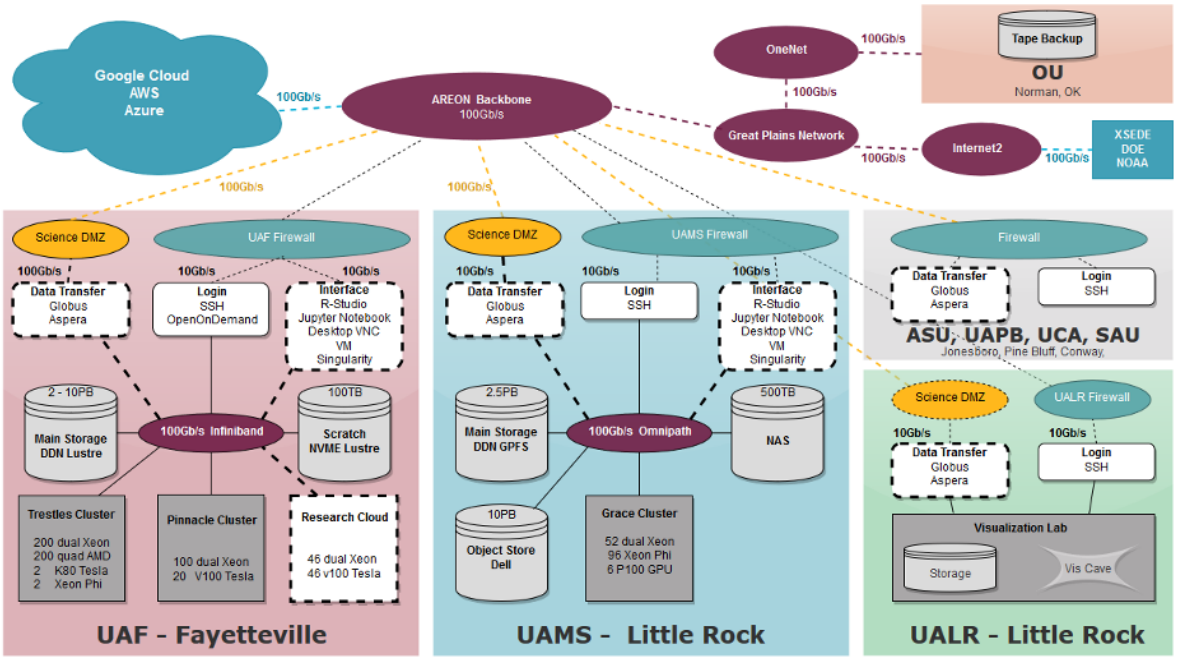


*Figure 2: The computational backbone of ARP consisting of existing and new, grant-supported equipment.*

ARP plans to connect Arkansas-based resources to the Open Science Grid (OSG) using gateway nodes, leveraging work being done for the CC* Compute award #2018766: GP-ARGO: The Great Plains Augmented Regional Gateway to the Open Science Grid. The connection will give DART and other researchers in Arkansas access to the vast resources for High Throughput Computing available on the Open Science Grid, while providing resources to the Open Science Grid that are not being used for local projects.

ARP will partner with The Carpentries to deliver high quality data science-oriented training in scientific software development, data management, and code management. UA is currently a Silver Member and will seek to offer at least 5 online training sessions per year in various locations.

### 3.2.4 *Protected and Sensitive Information Storage*

As noted in **recommendation 3**, the ability of institutions to manage protected research information under NIST 800.171 is quite limited. Based on early meetings of the CWG, efforts are underway at UAF and UAMS to develop and implement campus system security plans (SSP). These documents will identify the functions and features of a system, including all its hardware and the software installed on the system. They also define the security measures that have been or will be soon put in place to limit access to authorized users, as well as to train managers, users and systems administrators in the secure use of the system. They include details of processes for auditing and maintaining the system, in addition to information about how you plan to respond to security incidents that occur on the network. An SSP is a comprehensive summary of all security practices and policies that will help to keep CUI data secure if

the contractor is awarded a contract (typically DoD, DoE, but others as well). These two SPs will then be combined into an ARP SSP defining how and when sensitive data may be moved between sites.

We understand that this a critical element for ARP but given the level of coordination required to achieve immediate fulfillment is unlikely. Therefore, with assistance from CyberTeam and GPN, we will engage Trusted CI: The NSF Cybersecurity Center of Excellence. The added level of engagement with Trusted CI as well as with the IT departments at UAF, UAMS, and then other campuses, necessitates a change to the milestones in the DART Strategic. Plan. We propose to move a coordinated UAF/UAMS implementation of CUI management to Year 3 (Appendix B, Objective 1.1.e Activity 2).

3.2.5 ***Tiered CI Planning***

As a means of simplifying our approach to extensive federated identity requirements we have organized DART institutions (plus a number of other institutions) in integration tiers. Tier 1 designates institutions who have some measure of network architecture necessary to create Science DMZs and are active in DART. Tier 2 designates the remaining research institutions, while Tier 3 designates 2-year colleges whose IT departments who have express strong interest in connectivity to ARP (note that Pulaski Technical College and Northwest Arkansas Community College are the largest 2-year colleges in the state by enrollment).

*Table 3. ARCC Integration Tiers for ScienceDMZ inclusion. These same Tiers are defined in the CC* CIRA: SHARP proposal described in Appendix A.*

| Integration Tier | Entity | Contact |
|---|---|---|
| Tier 1 | UAF | Don DuRousseau |
| | UAMS | Michael Greer |
| | ASU | Henry Torres |
| | UALR | Thomas Bunton |
| Tier 2 | UCA | Trevor Seifert |
| | SAU | Mike Argo |
| | UAPB | Willette Totten |
| | UAM | Anissa Ross |
| | UA Cooperative Extension Service | Sam Boyster |
| Tier 3 | UA Pulaski Technical College | David Glover |
| | Northwest Arkansas Community College | Jason Degn |
| | UA Community College at Batesville | Steve Collins |

## 4. CI Learning Needs

***User Story:*** *A PhD student has a project that involves assessing changes in land classes across the continental United States, year-by-year, from 2008 – 2020. Each dataset includes more than 130 classes and more than 8 billion 30-meter pixels. The student is proficient in GIScience and performing analysis in R. Processing these data to year-to-year change would take months using traditional serial computing methods. While aware of high-performance computing capabilities the student no prior experience with the high-performance computing systems, job queueing, or troubleshooting that will enable more efficient analysis. The Open OnDemand portal provides intuitive access to high-performance computing resources for a broader campus audience and a "foot-in-the-door" for faculty, staff, and students who may have been previously deterred by unfamiliar environments and workflows. In this case, the student uses the Open OnDemand portal to: 1) work within a familiar GUI environment, RStudio, while building and testing workflows; 2) easily load required data, and download results, using the DTN via Open OnDemand; 3) use GitLab repositories to maintain and clone repositories; and 4) begin to develop an understanding of the SLURM job management system and code parallelization.*

This student's user story is typical of the nature of interaction we expect to see from DART researchers as well as general use of the ARP. It highlights the need for interactive data analysis and exploration, direct integration of one or more Git repositories directly into ARP clusters, the need to move large amounts of data to and from ARP, the need for parallelizing code (primarily R and Python).

The DART Strategic Plan defines a series of training workshops that will designed to meet the needs of student researchers and faculty whose computing needs differ from standard HPC clients. CI Research Theme objective 1.1.d, Activity 4 calls for leveraging the UAF Library's (contact person: Laura Lennertz) existing membership in The Carpentries to provide at least 5 workshops per year and train at least 2 instructors per year from non-UAF institutions. These workshops will focus on basic skills in Unix, Git, R and Python. These are meant to address deficiencies in new students or to expand the skillsets of experienced students. Activity 5 allows for more focused training on ARP resources. These will address direct integration of DART Gitlab repositories with Pinnacle and Grace, using Globus Basic to move large amounts of data to and from Pinnacle and Grace, and how to parallelize R and Python code on both clusters.

The need for a campus research computing champion structure (**recommendation 5**) could be met by on some campuses by The Carpentry-trained instructors and on others by a sub-group DART students and faculty. The nature of this engagement will be further addressed between the CI Research Theme and the GP CyberTeam.

# 5. Budget and Budget Justification

### 5.1.1 *CI Budget Summary*

*Table 4: Summary of CI direct costs (project-wide, by project year).*

| Organization | Yr 1 ($K) | Yr 2 ($K) | Yr 3 ($K) | Yr 4 ($K) | Yr 5 ($K) | 5-Yr Total | % of NSF Award | Cost Share ($K) |
|---|---|---|---|---|---|---|---|---|
| CI: Salaries and fringe benefits | 158 | 219 | 225 | 232 | 239 | 1,073 | 5.37 | 0 |
| CI: Equipment | 700 | 496 | | | | 1,196 | 5.98 | 0 |
| CI: Student support | 93 | 93 | 98 | 98 | 102 | 485 | 2.42 | 444 |
| CI: Faculty publication | 8 | 8 | 8 | 8 | 8 | 38 | 0.00 | 0 |
| CI: Faculty travel | 10 | 10 | 10 | 10 | 10 | 50 | 0.25 | 0 |
| Other: Globus Cloud | 40 | 40 | 40 | 40 | 40 | 200 | 1.00 | 0 |
| Total | 1,009 | 866 | 381 | 388 | 399 | 3,042 | 15.02 | 444 |

To support the research, the following equipment will be integrated per the award: 1) a dedicated data science-oriented cluster will be purchased and physically joined to Pinnacle at UAF. It will consist of approximately 46 nodes each with dual-Xeon Cascade Lade 20-core processors, 768 GB of memory, 480 GB local solid-state storage, one Nvidia Tesla V100 GPU. Nodes will be connected via EDR or HDR InfiniBand. 2) Additional storage (500TB) will be added at UAMS and 48 nodes from the existing Grace cluster re-tasked to contribute to ARP. To be investigated is the dynamic allocation of resources between ARP and traditional local HPC based on demand. Funds will be used to add 100Gb/s capability to the UAMS computing center in year 2 establishing a high-bandwidth connection between the two major computing clusters in the state via links provided by the statewide ARE-ON backbone.

### 5.1.2 *Proposed Budget Modifications*

The proposed budget modifications primarily address **recommendations 1 and 2**. Budget changes required to meet **recommendation 3** are unknown pending further CWG discussions but, if needed, would only supplement institutional funding at UAF and UAMS.

Given the need for a federated identity solution across at least the Tier 1 and 2 campuses recommended by the GP CyberTeam, we propose to reallocate a proportion of funds originally set aside for Globus Standard and Globus for Box subscriptions for seven campuses towards: 1) providing support to these same campuses for improving basic federated identify services and other network changes to be identified; and 2) support the purchase and setup of a dedicated server at UAF to run the two services that will be opened up to all DART project faculty, staff, and students: a) Open OnDemand web portal to AHPCC clusters and b) the Data Transfer Node (DTN). The DTN should be equipped with sufficient local storage to accommodate data transfers of 300+ users. Currently, Globus Cloud Services is being implemented using the Basic subscription; the availability of these services will not be impacted by the proposed budget modifications.

In both cases, these funds will remain in the EPSCoR Central Office budget and will be distributed to institutions as needs are identified during the ongoing meetings of the CWG. This change is justified by more fundamental needs that will have greater impact on the reach and useability of the ARP.

*Table 5. Summary of modified CI direct costs (project-wide, by project year)*

| Organization | Yr 1 ($K) | Yr 2 ($K) | Yr 3 ($K) | Yr 4 ($K) | Yr 5 ($K) | 5-Yr Total | % of NSF Award | Cost Share ($K) |
|---|---|---|---|---|---|---|---|---|
| CI: Salaries and fringe benefits | 158 | 219 | 225 | 232 | 239 | 1,073 | 5.37 | 0 |
| CI: Equipment | 727 | 496 | | | | 1,196 | 5.98 | 0 |
| CI: Student support | 93 | 93 | 98 | 98 | 102 | 485 | 2.42 | 444 |
| CI: Faculty publication | 8 | 8 | 8 | 8 | 8 | 38 | 0.00 | 0 |
| CI: Faculty travel | 10 | 10 | 10 | 10 | 10 | 50 | 0.25 | 0 |
| Other: Globus Cloud | | | | 40 | 40 | 80 | 0.50 | 0 |
| Other: Support for federated identity at key institutions | 13 | 40 | 40 | | | 93 | 0.50 | 0 |
| Total | 1,009 | 866 | 381 | 388 | 399 | 3,042 | 15.02 | 444 |

**Proposed Budget Modification 1: Federated Identity Support**

Funds budgeted in YR2 and YR3 ($80,000 total) for the Globus Standard subscription and Globus for Box subscription will be instead used during those years to support federated identify improvements, such as enrollment in InCommon, at various campuses that are needed to support the three major services being provided and promoted through the ARP.

**Proposed Budget Modification 2: Dedicated Server for Open OnDemand and DTN**

The following items will be purchased to complete the setup of this dedicated server; total cost will be approximately $27,000 (plus applicable tax and shipping). Funds previously budgeted for YR1 Globus Cloud Services ($40,000) will be used to support this purchase.

1) web portal/DTN node: PowerEdge 7525 with 2x 7543 and 768GB, 100Gb/s Ethernet card and cable (~ $9,000, plus tax and shipping);
2) JBOD storage enclosure ($3,055, plus tax and shipping); and,
3) 44 - 16TB (704TB RAW) drives ($335 x 44 = $14,750, plus tax and shipping).

# 6. Appendix A: CC* CIRA Proposal Development

As noted in Section 2: Gap Analysis and Needs Assessment, only two Universities in Arkansas have received CC* funding. The CI Research Theme is working through the CWG to stage a series of proposals over the next two years to coordinate network improvements and further leverage DART funding. The first proposal is targeted for the March 2021 solicitation due date and will be led by UAF Associate CIO for Research Don DuRousseau with Co-PI's from UAMS, UCA, and UALR.

Project Summary for CC* proposal to facility planning activities. DART is not dependent on this proposal but would be advanced by it.

**Overview:** The purpose of this CC* CIRA: Shared Arkansas Research Plan for Community Cyber Infrastructure (SHARP_CCI) proposal is to develop a statewide CI plan for Arkansas that focuses on eight (8) degree granting institutions performing Science and Engineering research on campuses across the state. Each school has a growing demand for federated access to high-speed networks, shared storage arrays, high-performance compute clusters, technical training and managed support services, and a coordinated plan for providing these capabilities and services does not currently exist. While considerable academic research is happening in Arkansas across the fields of Big Data Analytics, Genomics, Medical Science, Cybersecurity, Smart Farms and Distributed Sensing, some schools are better positioned than others to fully take advantage of existing compute resources located at UAF, UAMS, and UALR campuses. Thus, the goal of this proposal is to develop a coordinated CI plan to document the environments, capabilities, technology needs, and resource gaps at the 8 locations and design a statewide CI strategy to meet the current and future demands for a unified research cyberinfrastructure (RCI). In support of this effort, we will work closely with IT and research leaders at each school as well as with technical staff at ARE-ON, GPN, Open Science Grid (OSG), and the Engagement and Performance Operations Center (EPOC) to provide RCI engineering expertise, managed data services, and help in coordinating the CI plan implementation with our partnering schools. The completed statewide CI plan will ensure that each institution's RCI capabilities and needs are understood and the equipment, systems and services will be in place to provide easy and secure access to core RCI resources located throughout the state.

**Intellectual Merit:** The successful completion of this SHARP_CCI effort has the potential to advance S&E knowledge creation, distribution and utilization for academic, economic and social benefit in the state of Arkansas and beyond. Additionally, having a unified CI plan will lead to an economy-of-scale for not only building out a comprehensive RCI, but also for standing up a managed service environment where schools with limited resources and technical capability can receive assistance with data movement, access to storage and compute systems, and analytical processing and reporting support. The state-wide CI plan will include the instantiation of a Carpentries-based data science training program for researchers and students, and a managed RCI networking service in partnership with ARE-ON…as they are connected to all the schools throughout the state. In this manner, our state-wide CI plan development effort will provide the means for federated access to networks, compute, and storage systems across Arkansas, as well as provide access to education and training resources (i.e., S&E Courses, Coding Classes, Cyber Ranges, Internships) to support the operation, use and training of the core systems. To accomplish this, we will work with the IT leadership and research support groups at the 8 schools to create a Research Technology Support Team (RTST) to provide specialized expertise typically required for key research domains (e.g., Astrophysics, Bioinformatics, Communications, Cybersecurity, Medicine, Modeling & Simulation, ARC-GIS, ML/AI and NLP). The RTST is a main component of our CI plan and will be available to all researchers, educators and students across the state and other research communities beyond.

**Broader Impacts:** The potential of our SHARP_CCI project will benefit the Arkansas S&E research community as a whole and significantly contribute to the achievement of specific desired campus

outcomes at the 8 partnering schools located throughout the state (ASU, SAU, UAF, UACE, UALR, UAMS, UAPB, UCA). Our proposed SHARP_CCI effort will develop a coordinated state-wide campus research cyberinfrastructure (RCI) strategy that will lay out the technical foundations for each school and define the organizational commitment needed to advance knowledge, understanding, and education in support of many vastly different research programs. Most importantly, these diverse research programs across the state all share a single need for an integrated and enhanced RCI that provides easy access to local and remote data and meta-data repositories, libraries of investigative tools (e.g., for signal processing, data analytics, pattern classification, medical and spatial imaging and data visualization), computational systems and a broad community of practice in support of research and educational activities.

## 7. Appendix B: Modified Activity Table

*Table 6. Modified Activity Table: Highlights indicate updated objectives and goals since the Strategic Plan was submitted.*

| **Goal 1.1 (CI1)** | **Establish the Arkansas Research Platform as a shared data science resource across the jurisdiction** |
|---|---|
| **Objective 1.1.a:** Establish the Arkansas Research Computing Collaborative (ARCC)<br>**Objective 1.1.b:** Upgrade cluster for data science research activity and integrate with existing resources<br>**Objective 1.1.c:** Establish a science DMZ in Little Rock (UAMS, UALR) and high-speed connection with UAMS<br>**Objective 1.1.d:** Establish a data and code sharing environment (GitLab and Globus)<br>**Objective 1.1.e:** Develop a System Security Plan (UA) and to establish necessary controls to store and manage controlled unclassified, HIPAA-related, and proprietary information at both UA and UAMS (other institutions if possible) | |
| **Goal 1.1 Output Metrics**<br>**Hardware and software infrastructure improvements (5):**<br>-- Install, configure, and make available data science nodes on Pinnacle Portal<br>-- ScienceDMZ at UAF and UAMS/UALR<br>-- 100Gb connection between ScienceDMZ<br>-- Establish a dedicated DART GitLab repository<br>-- Setup Globus data management services to point at DART storage arrays<br>**Documentation and user guides (4):**<br>-- Create one (1) ARCC technical management document<br>-- Amend existing MOU for ARCC expansion<br>-- Create two (2) CI Plans (1 x UAF, 1 x UAMS)<br>-- Create one (1) GitLab user guidelines reference document<br>-- Create and implement a UA System Security Plan<br>**Workshops, demonstrations, and trainings (38):**<br>-- Host two (2) online workshops per year for onboarding to ARP resources in YR2-5 (8 total)<br>-- Host five (5) online software carpentry workshops per year in YR2-5 (20 total)<br>-- Train and certify two (2) new software carpentry instructors per year in YR2-5 (10 total)<br>**Applications and platforms (5):**<br>-- Create one (1) distributed computing testbed for HDFS, Apache Spark, others (DC)<br>--Create four (4) spatiotemporal testbeds for (CI/DC/SM/LP) | |

| **Objective 1.1.a** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1:** Create ARCC advisory board with regional partners (GPN) | Establish CI advisory board | | | | |
| **Activity 2**: Establish ARCC governance, operations, and staff between UA and UAMS | Create a document defining organizational structure, roles, and responsibilities of ARCC | | | | |
| **Activity 3**: Expand ARCC to include UALR as a provider and other DART participants as consumers | | Amend existing MOU for ARCC expansion | | | |
| **Activity 4**: Create UAF CI Plan to support DART (prior to 1.1.b and 1.1.c) | Create coordinated ARE-ON UAF, UAMS CI Plan | | | | |

| **Objective 1.1.b** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1**: Specify and purchase data science cluster based on document from 1.1.a | Issue UA purchase order for additional equipment | Receive data science nodes for Pinnacle (anticipated) | | | |
| **Activity 2:** Test and deploy hardware elements for Pinnacle expansion for DART | | Install, configure, and make available data science nodes on Pinnacle | | | |
| **Activity 3:** Install and configure data science cluster to work with existing resources at UA, UAMS, UALR resources | Collect testbed specifications and software/platform needs | Create containerized Hadoop-based testbed for DC | Create containerized testbeds for SA and SM | Create additional containerized testbeds for SA and SM | |

| **Objective 1.1.c** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1:** Specify and purchase 100Gb switch | | Issue UAMS purchase order for 100 Gb switch | Receive 100 Gb switch (anticipated) | | |
| **Activity 2:** Install 100Gb switch | | | Install and configure new 100 Gb switch | | |
| **Activity 3**: Establish ScienceDMZ at UAMS | Create UAMS CI Plan | Specify and acquire additional DMZ components | Establish and validate 100Gb link to UAF and integrated DMZ | | |

| **Objective 1.1.d** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1:** Create/identify federated identify or other authentication mechanism for all sites that provides access to core ARP resources | Establish federated ID for all project participants | | | | |
| **Activity 2**: Setup dedicated GitLab repository | Create and publish document outlining GitLab user guidelines, minimum standards for code repositories, and best practices. | | | | |
| **Activity 3:** Setup Globus Data Management Services | | Establish Globus Basic endpoints to DART storage arrays at key sites | | Globus Standard subscription executed | |
| **Activity 4:** Engage other research themes to develop research-specific training modules in e.g., Python, R, Git, HPC, Singularity | | -- Host 5 online software carpentry workshops<br>-- Train 2 software carpentry instructors | -- Host 5 online software carpentry workshops<br>-- Train 2 software carpentry instructors | -- Host 5 online software carpentry workshops<br>-- Train 2 software carpentry instructors | -- Host 5 online software carpentry workshops<br>-- Train 2 software carpentry instructors |

| | | | | | |
|---|---|---|---|---|---|
| **Activity 5:** Develop and deploy training materials for code sharing and large data transfer protocols | | Host 2 online ARP-specific training sessions | Host 2 online ARP-specific training sessions | Host 2 online ARP-specific training sessions | Host 2 online ARP-specific training sessions |
| | | | | | |
| **Objective 1.1.e** | **Specific Milestones** | | | | |
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1**: Identify the number and type (HIPAA, proprietary economic, CUI, etc.) of private and secure data sources that will need to be accessed by DART researchers. | Collect research theme needs | | | | |
| **Activity 2:** Setup capacity for storing and managing CUI and HIPAA data at UAF coordinated with UAMS | | Complete UA and UAMS System Security Plans | Combined SSP's into joint ARP SSP | | |

| Goal 1.2 (CI2) | Visualization for complex data in diverse data-analytics application domains |
|---|---|

**Objective 1.2.a:** Investigate state-of-the-art visualization solutions
**Objective 1.2.b:** Define domain-specific integration of visualization solutions
**Objective 1.2.c:** Introduce/integrate visualization for shared test beds

**Goal 1.2 Output Metrics**

**Publications, presentations, and reports (3):**

-- Three (3) presentations, reports, or other publications: 1 in YR1 and 2 in YR2

**Workshops, demonstrations, and trainings (4):**

-- One (1) online workshops per year for advanced visualization in YR2-5 (4 total)

**Applications and platforms (9):**

-- Develop one (1) visualization solution for each research theme, including CI (5 total)
-- Integrate one (1) visualization into existing testbed for each research theme (4 total)

| **Objective 1.2.a** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1**: Investigate/define state-of-the-art visualization | 1 presentation/report | | | | |
| **Activity 2:** Investigate standard tools for data science visualization | | 1 presentation/report | | | |
| **Activity 3:** Investigate/define data exchange strategies and their relationship to other research themes | | 1 presentation/report | | | |

| **Objective 1.2.b** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1:** Develop and deploy visualization infrastructure software | Collect research theme needs | 1 software prototype (alpha) | 1 software prototype (beta) | | |
| **Activity 2:** Develop domain-specific visualization solution for DC | | | 1 software prototype (alpha) | 1 software prototype (beta) | |

| | | | | | |
|---|---|---|---|---|---|
| **Activity 3:** Develop domain-specific visualization solution for SA | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 4:** Develop domain-specific visualization solution for SM | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 5:** Develop domain-specific visualization solution for LP | | | 1 software prototype (alpha) | 1 software prototype (beta) | |

| **Objective 1.2.c** | **Specific Milestones** | | | | |
|---|---|---|---|---|---|
| | Year 1 | Year 2 | Year 3 | Year 4 | Year 5 |
| **Activity 1:** Integrate visualization into existing testbeds for automated data curation environment DC/SM | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 2:** Integrate visualization into existing testbeds for social media-linked, GIS platform CI/DC/SM/LP | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 3:** Integrate visualization into existing testbeds for bioinformatics workflows DC/SM | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 4:** Integrate visualization into existing transaction-based testbed LP/DC | | | 1 software prototype (alpha) | 1 software prototype (beta) | |
| **Activity 5:** Engage other research themes to develop research-specific advanced visualization training | | Host 1 online advanced visualization workshops | Host 1 online advanced visualization workshops | Host 1 online advanced visualization workshops | Host 1 online advanced visualization workshops |

# 8. Appendix C: Modified Cyberinfrastructure Logic Model

*Table 7. Modified Logic Model: Highlights indicate updated inputs, objectives, outputs, or outcomes since the Strategic Plan was submitted.*

| Research Theme | Input(s) | Objectives | Output(s) | Short-term Outcomes | Medium Term Outcomes |
|---|---|---|---|---|---|
| **Research Theme 1: Cyber Infrastructure** | **Goal 1 (CI1)**<br><br>**Staff**: Cothren, Prior, Springer, Chaffin, Tarbox, Deaton, DuRousseau, Pummill, Merrifield<br>**Partnerships**: Great Plains Network | Objective 1.1.a: Establish the Arkansas Research Computing Collaborative (ARCC)<br><br>Objective 1.1.b: Upgrade cluster for data science research activity and integrate with existing resources | **Hardware and Software Infrastructure:**<br>-- Install, configure, and make available data science nodes on Pinnacle Portal<br>-- ScienceDMZ at UA and UAMS/UALR<br>-- 100Gb connection between ScienceDMZs.<br>-- Establish dedicated DART GitLab repository<br>-- Setup Globus data management services to point at DART storage arrays<br><br>**Documentation and User Guides:**<br>-- Create a technical management document defining organizational | -- 100% increase in active accounts on Pinnacle and Grace<br>-- 100% increase in overall use measured in gigaflops<br>-- Code developed by DART researchers is shared via GitLab repository linked to public GitHub<br>-- Sharing of large data sets among individuals, institutions, and HPC clusters. | -- enhanced academic collaboration across Arkansas campuses measured by increased publications using CI resources<br>-- enhanced collaboration between industry and academia measured by the number of such projects that use CI resources<br>-- trusted data sharing between collaborating partners measured by the number of recorded data transfers and the total number of TB transferred |

| Research Theme | Input(s) | Objectives | Output(s) | Short-term Outcomes | Medium Term Outcomes |
|---|---|---|---|---|---|
| | | Objective 1.1.c: Establish a Little Rock (UAMS, UALR) ScienceDMZ and high-speed connection with UAMS | structure, roles, and responsibilities of ARCC for personnel at participating campuses<br>-- Amend existing MOU for ARCC expansion<br>-- UAF and UAMS will create CI Plans to support DART (1 x UAF, 1 x UAMS); these CI Plans will serve as templates for other Institutions<br>-- Create and publish document outlining GitLab user guidelines and minimum standard for code repository<br>-- Create System Security Plan at UA<br><br>**Workshops, demonstrations, and trainings:**<br>-- Two (2) online workshops per year for onboarding to ARP resources in YR2-5<br>-- Five (5) online software carpentry workshops per year in YR2-5 focusing on | | |
| | | Objective 1.1.d: Establish a data and code sharing environment (GitLab and Globus) | | | |

| Research Theme | Input(s) | Objectives | Output(s) | Short-term Outcomes | Medium Term Outcomes |
|---|---|---|---|---|---|
| | | Objective 1.1.e: Establish necessary controls to store and manage controlled unclassified, HIPAA-related, and proprietary information at UA and UAMS (other institutions if possible) | developing and sharing code and data; data science programming; and data management<br>-- Train and certify two (2) new software carpentry instructors (across the jurisdiction) per year in YR2-5<br><br>**Applications and platforms:**<br>-- Create one (1) distributed computing testbed for HDFS, Apache Spark, others (DC)<br>- -Create four (4) spatiotemporal testbeds for (CI/DC/SM/LP) | | |
| | **Goal 2 (CI2)**<br><br>**Staff:** Springer, Conde, Huff, Milanova<br>**Equipment:** As determined by need and existing capability | Objective 1.2.a: Investigate state-of-the-art visualization solutions | **Workshops, demonstrations, and trainings:**<br>-- One (1) online workshops per year for advanced visualization in YR2-5 | | |

| Research Theme | Input(s) | Objectives | Output(s) | Short-term Outcomes | Medium Term Outcomes |
|---|---|---|---|---|---|
| | | Objective 1.2.b: Define domain-specific integration of visualization solutions | **Publications, presentations, and reports:**<br>-- Three (3) presentations, reports, or other publications: 1 in YR1 and 2 in YR 2<br><br>**Applications and platforms:**<br>-- Develop one (1) visualization solution for each research theme, including CI (5 total)<br>-- Integrate one (1) visualization into existing testbed for each research theme (4 total) | | |
| | | Objective 1.2.c: Introduce/integrate visualization for shared test beds | | | |
| | | Objective 3.7.b: Explore how to protect the privacy of classification input data from the server hosting machine learning models | | | |
| | | Objective 3.7.c: Assess/Protect the trustworthiness of training data and machine learning models | | | |

## 9. Appendix D: Comprehensive "State of the Network" Across the Resource Providers