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

External Evaluation Report

July 1, 2020 to February 28, 2023

Prepared for:
Arkansas Economic Development Commission

Author:
Kirk F. Minnick

Minnick & Associates, Inc.
PO Box 820 Tijeras, New Mexico 87059
505-889-9358 Fax: 505-212-5842
Kminnick@EvalTeam.com

This document is supported by federal funds from the National Science Foundation (NSF), Office of Integrative Activities (OIA), contract number (OIA-1946391). Its contents do not necessarily reflect the views or policies of NSF, and no official endorsement of material should be inferred.

Introduction

Arkansas Data Analytics that are Robust and Trusted (DART) is funded by the National Science Foundation (NSF) and is designed to help fulfill the foundation's mandate to promote scientific progress nationwide. The EPSCoR program is directed at those jurisdictions that have historically received lesser amounts of NSF Research and Development (R&D) funding. Twenty-five states, the Commonwealth of Puerto Rico, Guam, and the U. S. Virgin Islands are currently eligible to participate. NSF establishes partnerships with government, higher education and industry that are designed to effect lasting improvements in a state's research infrastructure, R&D capacity and its national R&D competitiveness.

DART was awarded funding of $20 million for five years beginning July 1, 2020. The overarching goals and focus are outlined in the proposal and reiterated below:
DART will develop:

1) the means to increase the speed and efficiency of data curation and labeling,
2) techniques to protect privacy and impartial content,
3) methods for harnessing the predictive power of machine learning while increasing the interpretability of the processes behind the predictions, and
4) data science curricula that are more inclusive and better prepare students for a data-centric future.

The project is organized into nine components:

1. Coordinated Data Science Infrastructure (CI)
2. Data Life Cycle and Curation (DC)
3. Social Awareness (SA)
4. Social Media and Networks (SM)
5. Learning and Prediction (LP)
6. Education (ED)
7. Workforce Development (WD)
8. Communications & Dissemination Component (CD)
9. Broadening Participation (BP).

Evaluation Process

Data for evaluation metrics and for reports to NSF-EPSCoR are collected electronically. DART has implemented the EPSCoR Reporting Core (ER-Core) which provides secure access to project participants for reporting and assessment.

Both the quantitative and qualitative data are collected using a variety of methodologies. The formative evaluation is focused on component development and the process of implementation, including feedback from students, faculty/researchers and our observations. The Evaluation Team collects and evaluates formative data to assist project leadership in assuring quality of program management and effective project development and implementation. Data from meeting minutes, project documentation, interviews and observations, surveys and participant feedback help inform the formative evaluation.

The program leadership, External Evaluator, External Advisory Board (EAB) and Industry Advisory Board (IAB) monitor and assess program activities. Project leadership is responsible for ensuring that data is collected on milestones, participants, proposals/awards, publications, and other project outputs, as well as implementation of recommendations from the EAB, IAB and evaluator. The evaluator develops and administers surveys, observes project activities, analyze data, and produce evaluation reports. The EAB and IAB assess overall results and progress towards objectives and identify problem areas.

The evaluation is designed to answer the following progress and outcomes questions:

- Are DART researchers becoming more competitive for R&D funding?
- Is the DART research generating knowledge that is being disseminated and applied in academia, industry and government?
- Are state and regional collaborations being fostered that promote research, innovation, and benefit society?
- Is the DART broadening participation of its people (especially those historically underrepresented), institutions and organizations in STEM?
- Is Arkansas capitalizing on the investment to further develop experimental programs in data analytics?

Evaluation and Data Collection Methods

Data for the evaluation comes from a variety of sources and are collected using a variety of methodologies, including both formative and summative evaluation strategies. The following provides a sample of the data sources and methodologies that are used.

Table I: Evaluation Data Sources

Data Source	Description	Purpose
Feedback Surveys	Participants in DART activities/events will be asked to provide feedback on the activity in which they were involved. Most surveys will be administered online through email invites to the participant.	These provide "customer satisfaction" feedback for project activities, help the project identify participant needs for future events/activities and assess the efficacy of materials and displays.
Observations	The evaluator will observe and participate in many of the project activities. These will include attending student outreach and faculty workshops, All Hands Meetings and project meetings.	These observations help provide context for participant survey feedback and can help activity leaders make real time changes during the activity.
Project Documents	Project reports and other written documents, such as the strategic plan; provide the evaluator with the research plans and accomplishments.	These documents provide the plans and accomplishments in a more holistic manner, which can be compared to the reported project data.
Project Data (ERCore)	Project data on participants, proposals, awards, presentations, publications, collaborations, etc. are collected by DART through the ERCore portal.	These data provide some of the basic outputs and outcome measures used in tracking summative results and reviewed for consistency and reliability.
Group/individual interviews	Interviews will be conducted with individuals and groups of faculty to gather feedback on the project.	These interviews provide a perspective on the project from the field.
NSF Awards Database	This database provides the proposals funded in AR, along with the funding amounts and duration. It also contains the names of the PI and Co-PIs.	These data provide the data to assess the success of Arkansas researchers in competing for NSF funding and track AR EPSCoR faculty and student success.
Web of Science: Citation database	Web of Science provides access to a database of scientific publications and citations that can be searched by individual author, institution, or state	These data provide the number/citations of publications by participants. These data can be examined by time period as well as by eywords.

Formative/Implementation Evaluation

The formative or implementation evaluation examines how the project proposed activities are being implemented. The evaluators examine how the activities are conducted; including the timing, participants involved, location, venue, content, project resources utilized, alignment with proposal, as well as the immediate outputs from those activities. This aspect of evaluation is designed to help improve future project activities by soliciting feedback from participants, conducting observations, and assessing the alignment between the proposed activities and implementation. Process evaluation metrics typically involve number and diversity of participants; participation satisfaction; number of proposals, presentations and papers submitted; number of collaborations; amount of equipment purchased and when; as well as other immediate outputs resulting from project activities. Formative evaluation feedback to project leadership is designed to help improve future activities and identify any discrepancies between proposed activities and timelines and implementation, including the involvement of the appropriate participants. The purpose of the formative evaluation is to help the project have the best implementation of its planned activities, so that it has the best opportunity to achieve its stated goals.

Table II
Formative/Implementation Evaluation

Evaluation Question	Data Sources	Frequency
Are the individual components being implemented as planned?	ER Core Annual Report	Annual
Are the appropriate staff/faculty/partners involved and working together towards the component goals(s)?	ER Core Annual Report	Annual
Are there adequate resources/materials/equipment available?	ER Core Annual Report	Annual
Are the appropriate participants selected and involved in activities/programs?	ER Core Annual Report	Annual
Do the activities/strategies match those described in the strategic plan/proposal? If not, are the changes in activities justified and described?	ER Core Annual Report	Annual
Are activities being conducted according to the proposed timelines? By the appropriate personnel?	ER Core Annual Report	Annual
Do project participants report high satisfaction in their participation in the activities?	Faculty and Graduate Student Survey	Annual

Progress/Summative Evaluation

The progress/summative evaluation assesses data that are collected from participants through a secure web-based portal, including data on participant numbers and demographics, collaborations (individual and institutions), new software tools, methodologies, presentations, publications, proposals, awards, and patents. These data are used to track outputs and outcomes for faculty, students, and the project, and are analyzed by year and cumulatively to ensure that the project is on-track to achieve annual and final project targets. The summative evaluation assesses key progress and outcomes questions, such as:

- Are the research components reducing the barriers of big data management, security and privacy and model interpretability and expanding the use of data analytics in industry and government?
- Are collaborations being expanded among people, institutions, and disciplines within Arkansas, as well as outside the state?
- Is the project broadening participation of people, institutions, and organizations in STEM with specific emphasis on data analytics? and finally,
- Are the workforce activities producing a diverse group of next generation data scientists that can apply these new technologies in academia, industry, and government?

People (Human Infrastructure, Collaboration and Diversity)

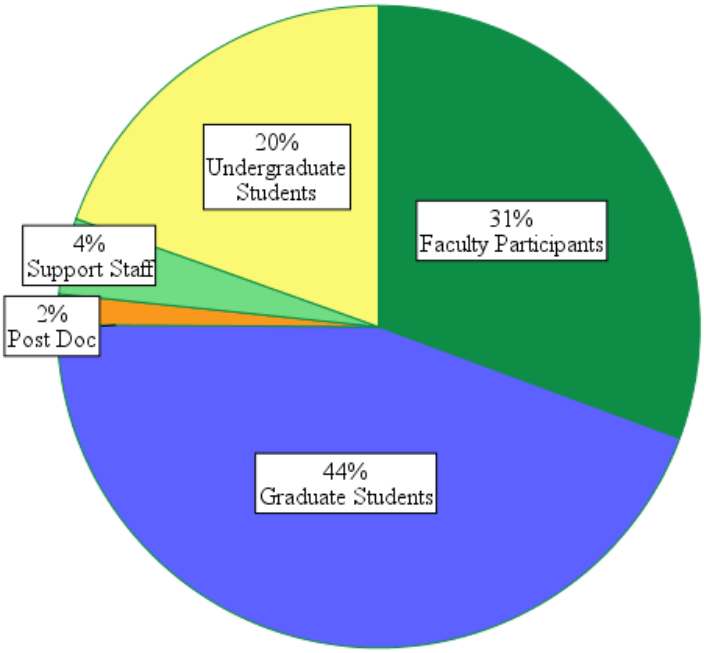
Diversity and expanding the involvement of women and underrepresented minorities in the sciences, is one of the objectives of NSF, EPSCoR and DART. This section explores the degree of involvement of women and underrepresented minorities by level of participation, participant role and participating institutions in the initiative. The role of collaborators and the nature of their collaborations with the project are also examined.

In this report, there are two types of participants: direct and indirect. There are direct project participants, those that have participated in a significant way in accomplishing the goals of DART; either in one of the research areas or education outreach/external engagement. Direct participants are all individuals who have been individually identified. Indirect project participants are the students of the K-12 teachers who have received professional development, members of the public or education institutions who have participated in outreach events. These latter participants are reported in the external engagement section.

Participants

DART has involved 261 direct participants in the project so far. Figure 1 shows the percentage of participants by their role.

Figure 1
Participant Roles
(N=261)



Almost half (44%) of the project participants are ‘Graduate students’ while another one-fifth (20%) are 'Undergraduates’. Less than one-third (31%) are ‘Faculty participants’ with Post Docs representing 1% of the total number of participants. The remaining participants were 'Support staff’ (4%).

It is important to examine the ethnic and gender diversity of participants by their role, as well, to ensure that diversity goals are being addressed across the different levels of participants. Figure 2 shows the number and percentage of under-represented ethnic minorities and females by participant role for all program years, as well as the unduplicated count of participants across all years. Demographic data are self-reported and do not represent all participants.

Figure 2
Participants by Role and URM* and Female*
By Program Year

Role	Year 1 Baseline	Year 2	Year 3	Unduplicated Count
Faculty	56 (41%)	75 (35%)	80 (31%)	86 (31%)
URM-Ethnic	2 (4%)	2 (3%)	3 (4%)	3 (4%)
Female	16 (29%)	22 (29%)	23 (29%)	24 (29%)
Post docs	1 (1%)	3 (1%)	4 (1%)	4 (1%)
URM-Ethnic	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Female	1 (100%)	2 (67%)	3 (75%)	3 (75%)
Graduate Students	47 (34%)	87 (40%)	116 (44%)	118 (42%)
URM-Ethnic	5 (13%)	9 (12%)	12 (12%)	13 (13%)
Female	18 (38%)	26 (30%)	39 (34%)	40 (35%)
Undergraduates	25 (18%)	41 (19%)	51 (20%)	62 (22%)
URM-Ethnic	9 (36%)	18 (45%)	23 (46%)	28 (46%)
Female	12 (48%)	19 (46%)	23 (45%)	29 (48%)
Support Staff	8 (5%)	10 (5%)	9 (4%)	12 (4%)
URM-Ethnic	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Female	4 (50%)	5 (50%)	6 (60%)	6 (55%)
Total	137	216	261	282
URM-Ethnic	16 (13%)	29 (15%)	38 (16%)	44 (17%)
Female	51 (37%)	74 (34%)	94 (36%)	102 (37%)

* URM and Female percentages are based on the subset of participants who self-reported their ethnicity and gender

DART involved 261 participants in program Year 3, an increase of 45 (21%) from Year 2. In Year 3, undergraduate students continue to be the most ethnically diverse group of participants with 46% self-identifying as an ethnic minority, while 12% of graduate students and 4% of faculty self-identify as an ethnic minority. No post-docs or support staff self-identified as an ethnic minority. The overall ethnic diversity across all participant groups was 16%.

Female diversity is much higher with more than one-third (36%) of all participants self-reported being female. The participant groups with the most females were post-docs (75%); support staff (60%); undergraduates (45%); graduate students (34%) and faculty (29%).

Over the three program years DART has involved 282 unique individuals: almost one-third (31%) faculty; over one-third (42%) graduate students and more than one-fifth (22%) undergraduates. Across all participant groups 17% self-identified as an ethnic minority and one-third (37%) self-identified as female.

Figure 3 presents the number and percentage of participants by institution by program year and an unduplicated count across the program years.

Figure 3

Number/Percent of Participants* by Institution and Program Year

Institution	Year 1	Year 2	Year 3	Unduplicated Count
Arkansas Economic Development Commission	5 (4%)	9 (4%)	6 (2%)	8 (3%)
Arkansas State University	7 (5%)	9 (4%)	12 (5%)	12 (4%)
Arkansas Tech University	2 (2%)	11 (5%)	8 (3%)	13 (5%)
North Arkansas College	1 (1%)	1 (1%)	1 (<1%)	1 (<1%)
Philander Smith College	3 (2%)	7 (3%)	13 (5%)	14 (5%)
SAU Tech	0 (0%)	0 (0%)	1 (<1%)	1 (<1%)
Shorter College	4 (3%)	4 (2%)	4 (2%)	5 (2%)
Southern Arkansas University	11 (8%)	14 (7%)	17 (7%)	19 (7%)
University of Arkansas at Fayetteville	50 (37%)	63 (29%)	83 (32%)	86 (31%)
University of Arkansas at Little Rock	30 (22%)	67 (31%)	77 (30%)	80 (28%)
University of Arkansas at Pine Bluff	5 (4%)	4 (2%)	7 (3%)	8 (3%)
University of Arkansas for Medical Sciences	11 (8%)	16 (7%)	21 (8%)	21 (7%)
University of Central Arkansas	8 (6%)	11 (5%)	11 (4%)	14 (5%)
All Institutions	**137**	**216**	**261**	**282**

* does not include advisory board members

The institutions with the largest number of participants involved all project years are: University of Arkansas-Fayetteville (31%), University of Arkansas at Little Rock (28%); University of Arkansas for Medical Sciences (7%); Southern Arkansas University (7%); Arkansas Tech University (5%) and University of Central Arkansas (5%). The project has also involved participants from other higher education institutions in the state, including many smaller institutions, such as Arkansas State University (4%); Philander Smith College (5%); University of Arkansas at Pine Bluff (3%); Shorter College (2%) and North Arkansas College (<1%).

Collaborations

Recognizing that for DART to be successful at improving its research competitiveness, it must maximize the collaborations between researchers within and outside the state. Figure 4 presents the number of DART collaborators by institutional type and location.

Figure 4
Number of External Collaborators by Type of Institution and Location by Program Year

Type of Institution	Year 1	Year 2	Year 3	Unduplicated
Academic-Research Institution	4	8	12	12
Academic-Primarily Undergraduate	3	2	2	4
Industry/Business	2	4	3	4
Non-Profit	1	1	1	1
K-12 School/Provider	0	2	2	2
State Agency	0	2	2	2
Other	0	3	3	3
Location of Collaborator	**Year 1**	**Year 2**	**Year 3**	**Unduplicated**
Within Arkansas	5	16	18	18
Outside Arkansas, but in US	4	5	5	8
International	1	1	2	2
Total Collaborators	10	22	25	28

Project participants reported 28 external collaborators during the first three project years. More than one-third (43%) of the collaborators are at 'academic research institutions', while more than one-tenth (14%) are from 'primarily undergraduate institutions' and 'industry' (14%). Almost two-thirds (64%) of the external

collaborators reported are in Arkansas; while one-fourth (29%) are collaborators located outside Arkansas but within in the US. Only 7% of the external collaborators are from outside the US.

While it is important to increase the number of collaborations with other researchers, it is equally important to involve collaborators from a variety of institutions. Figure 5 presents the number of institutions in which the DART collaborators work by institutional type and location.

Figure 5
Number of Collaboratoring Institutions by Type of Institution and Location by Program Year

Type of Institution	Year 1	Year 2	Year 3	Unduplicated
Academic-Research Institution	2	5	9	9
Academic-Primarily Undergraduate	3	2	3	4
Industry/Business	2	4	3	5
Non Profit	1	1	1	1
K-12 School/Provider	0	2	2	2
State Agency	0	1	1	1
Other	0	3	3	3
Location of Institution	**Year 1**	**Year 2**	**Year 3**	**Unduplicated**
Within Arkansas	3	12	14	14
Outside Arkansas, but in US	4	5	5	9
International	1	1	2	2
Total Collaborating Institutions	8	18	21	25

DART external collaborators are at 25 different institutions; one-third (36%) are 'Research Institutions', while one-fifth (20%) work at 'Industry/Business' and less than one-fifth (16%) are at 'Primarily Undergraduate Institutions'. Over half (56%) of the external collaborating institutions are in Arkansas.

Material Infrastructure (Equipment, Models and Cyberinfrastructure)
As the name implies, the Research Infrastructure Improvement (RII) funding is intended to provide the resources to enhance the research capabilities of the jurisdiction and become more competitive by acquiring equipment and funding infrastructure necessary for world-class research.

Over $1.2 million has been expended or ordered during the project to purchase computer infrastructure at three institutions. In Year 1, project funds were used to purchase $650k of infrastructure at the University of Arkansas at Fayetteville. The equipment purchased included DELL Fiber Splitter cables, PowerEdge XE8545, Power Edge R7525, Server, NVIDIA Ampere A100 649,607.18 and a 40-port Mellanox Quantum QM8790. This equipment is part of the DART CI Plan that increases additional hardware needed to move pinnacle out from behind the firewall. Also, in Year 1 data storage servers allowing data sharing among DART researchers were purchased for $24k at the University of Arkansas at Little Rock.

In Year 2, $496k was used to upgrade the research backbone at University of Arkansas Medical Sciences to collaborate with ARE-ON and extend service to the University of Arkansas at Fayetteville.

In Year 3, no project funds were budgeted or used to purchase equipment or infrastructure.

Research (Observing, Data Collecting, Discovery, Funding Support)
Data collected, observations or field work, research conducted, and proposals submitted and awarded provide an indicator of the research outputs accomplished by the efforts of the DART researchers. Figure 6 lists the number of proposals and amount of funding requested by funding source since DART started.

Figure 6
Proposal Success Rates and Amounts Proposed, Funded and Awarded as of February 29, 2023

Funding Source	Number Proposed	Amount Proposed	Number Pending	Amount Pending	Number Funded	Amount Funded
Arkansas Biosciences Institute	2	$300,000	0	$0	1	$150,000
Arkansas INBRE	1	$328,389	0	$0	0	$0
Arkansas NASA EPSCoR	1	$40,000	1	$40,000	0	$0
Arkansas Research Alliance	2	$175,000	0	$0	2	$175,000
NASA	2	$244,905	1	$99,999	1	$144,906
National Institutes of Health (NIH)	12	$19,207,485	7	$15,544,957	3	$2,841,526
National Science Foundation (NSF)	33	$36,588,987	5	$2,798,093	16	$8,774,935
National Security Agency (NSA)	2	$26,333	1	$8,333	1	$18,000
Private Companies	4	$184,351	0	$0	3	$137,340
US Air Force: Office of Scientific Research (AFOSR)	2	$1,529,889	0	$0	2	$1,529,889
University Faculty Research Grants	11	$686,128	0	$0	10	$586,128
US Census Bureau	1	$714,198	0	$0	1	$714,198
US Department of Agriculture (USDA)	3	$1,392,976	1	$297,976	2	$1,095,000
US Department of Defense (DoD)	6	$12,281,115	0	$0	4	$7,671,214
US Department of Energy	5	$6,311,569	1	$1,997,797	1	$800,000
US Department of Health & Human Services	3	$2,661,707	0	$0	1	$2,461,707
US Department of Transportation	1	$155,199	0	$0	1	$155,199
US Geological Surveys (USGS)	1	$25,000	0	$0	1	$25,000
US Office of Naval Research (ONR)	4	$4,463,091	0	$0	4	$4,463,091
Overall	**96**	**$87,316,322**	**17**	**$20,787,155**	**54**	**$31,743,133**

Ninety-six proposals requesting over $87 million have been submitted by DART participants. As of February 29, 2023, 54 proposals have been funded for a total of $31.7 million, while seventeen are still pending a funding decision. The funding agencies where most of the proposals have been submitted are NSF (34%) and NIH (13%), while the most award dollars have come from NSF ($8.7M or 27%); US Department of Defense ($7.7M or 24%) and US Office of Naval Research ($4.5M or 14%).

Figure 7 presents a list of those awards over $500,000.

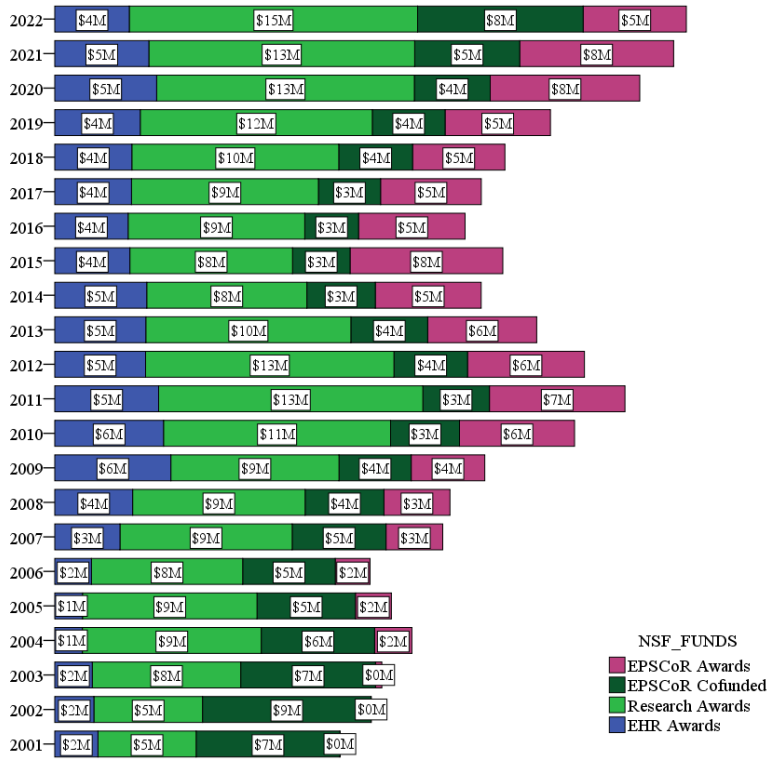
Figure 7
Awards over $500,000

Title of Award	Principal Investigator	Award Amount	Funding Source
Multi-Level Models of Covert Online Information Campaigns	Nitin Agarwal	$4,965,214	U.S. Department of Defense (DoD)
Developing Rapid Response Capabilities to Evaluate Emerging Social Cyber Threats	Nitin Agarwal	$3,773,526	U.S. Office of Naval Research (ONR)
Fusing Narrative and Social Cyber Forensics to Understand Covert Influence	Nitin Agarwal	$2,500,000	U.S. Department of Defense (DoD)
Targeting heat shock protein 72 to improve renal function after transplantation	Se-Ran Jun	$2,461,707	National Institutes of Diabetics and Digestive and Kidney Diseases
Center for studies of host response to cancer therapy	Se-Ran Jun	$2,280,000	NIH
RII Track-2 FEC: Artificial Intelligence on Sustainable Energy Infrastructure Network (AI SUSTEIN) and Beyond towards Industries of the Future	Haitao Liao, Xintao Wu, Xiao Liu	$1,450,003	NSF
Assessment of antibiotic resistance in fresh vegetables from farm to fork	Se-Ran Jun	$1,000,000	USDA
Statistics-Machine Learning-Enhanced (SMILE) Accelerated Reliability Growth of Complex Systems Based on Computer Experiments and Physical Tests	Haitao Liao	$869,482	Office of the Secretary of Defense through Air Force Institute of Technology
Photogrammetry Services, Task Order for CY2021 View Text	Jackson Cothren, Chase Rainwater	$800,000	Department of Energy
Next Generation Entity Resolution	John Talburt, Xiaowei Xu, Mariofanna Milanova	$714,198	U.S. Census Bureau
FAI: A novel paradigm for fairness-aware deep learning models on data streams	Xintao Wu	$628,789	NSF
IUCRC Phase I The University of Arkansas: Center for Infrastructure Trustworthiness in Energy Systems (CITES)	Qinghua Li	$525,000	NSF

While tracking the funding levels of DART researchers is critical to measuring the success of this award, it is equally important to track the NSF award dollars received by all researchers in the state.

Figure 8 presents the dollar amount converted to 2022 inflation adjusted dollars of NSF awards to Arkansas for the last ten years. The award dollars are distributed over the projected length of each award to smooth out the effect of a large award causing a spike in an individual year. This technique provides a better representation of when the dollars are expended and makes it easier to track trends in funding levels, now and in the future.

Figure 8
NSF Funding Awarded to Arkansas
NSF Funded Awards in Arkansas by Calendar Year: 2001 to 2022
Awarded Funding Equally Distributed Across Project Years
(Reported in 2022 Inflation Adjusted Dollars)



Note: 2010-2013 were American Recovery and Reinvestment Act years;
while 2020 and 2021 saw increased funding from Covid-19 Stimulus funding

NSF funding in the state has increased from $14 million in 2001 to $32 million in 2022, in 2022 adjusted dollars. While the years 2010 to 2013 showed marked increases in award dollars to the state, this was the result of the American Recovery and Reinvestment Act. A similar increase in Federal funding occurred in 2020 and 2021 with the Covid-19 Stimulus funding. However, Arkansas is on track to maintain a high level of award dollars coming into the state.

Since 2001, Arkansas has received over $493 million dollars in NSF funding. Most (43%) of this funding has been in the form of research awards totally $215 million; co-funded awards (both for research and EHR) have amounted to $102 million (21%); EPSCoR awards (Tracks1, 2 and 3) have amounted to $94 million (19%) and $82 million (17%) has been awarded for EHR proposals.

NSF funding for Research, excluding EPSCoR co-funding and Track1, has increased from $5 million in FY 2001 to over $15 million in FY 2022, an increase of 200%. In addition, Education and Human Resource funding from NSF has doubled from $2 million in FY 2001 to $4 million in FY 2020.

Knowledge Generation (Professional Presentations, Publications, Patents)
Professional presentations, posters and invited talks are critical to increasing the visibility and reputations of DART researchers, in addition to disseminating their valuable research findings to their colleagues. Figure 9 presents the number and percent of presentations made by presentation type and program year.

Figure 9
Presentations by Type of Presentations and Program Year

Type of Presentation	Year 1	Year 2	Year 3	Total
Invited Speaker	18	15	15	**47 (33%)**
Panel	1	0	1	**2 (1%)**
Poster	1	10	18	**30 (21%)**
Presentation/Talk	22	20	20	**62 (43%)**
Other	0	0	2	**2 (1%)**
Totals	**42**	**45**	**56**	**143**

DART participants reported making 143 presentations, posters, and invited talks during the first three program years. Almost half (43%) were 'Presentation/Talks', while one-third (33%) were 'Invited Speakers' and one-fifth (21%) were 'Posters'.

Publications

Dissemination of the knowledge and research findings by DART researchers is also an important outcome of the project. Figure 10 lists the number of publications by level of support from DART and program year.

Figure 10
Publications by Level of Support and Program Year

Level of Support	Year 1	Year 2	Year 3	Totals
Partial	6	70	30	106
Primary	2	19	16	37
Totals	**8**	**89**	**46**	**143**

Researchers reported 143 peer reviewed publications in the first three years of the project. Over one-fourth (26%) reported receiving primary support from DART. Some of the journals in which DART researchers have published included: BMC Bioinformatics; Infection and Immunity; Metabolites; Microbial Genomics and International Journal of Advanced Computer Science and Applications to name a few.

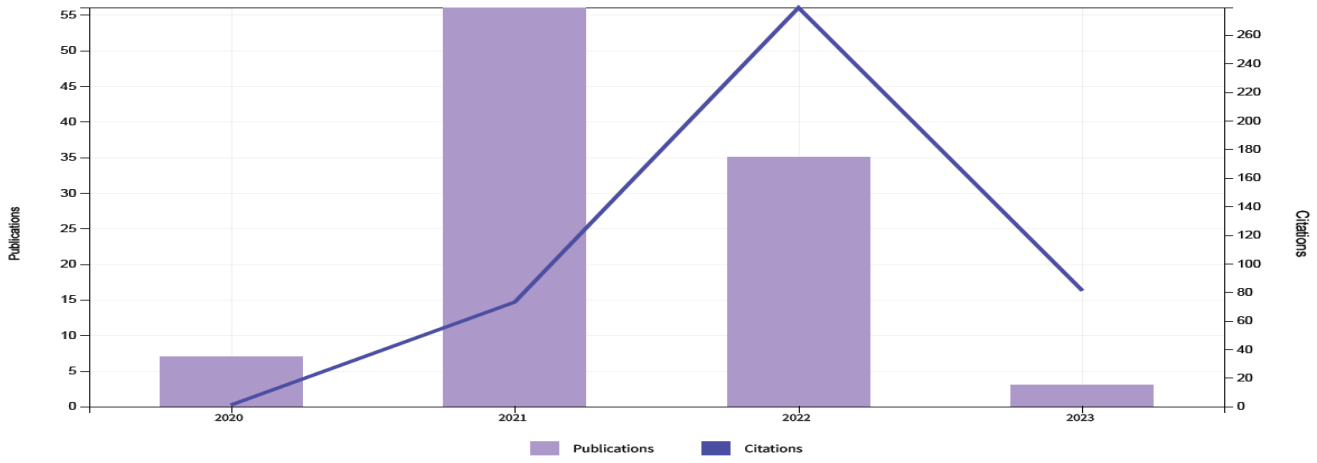
An analysis of these peer reviewed publications from the first three years was conducted using Clarivate's Web of Science. A total of 101 or 71% of the 143 publications were able to be matched with those in the citation index. Figure 11 presents a chart from the Web of Science that groups the publications into the content areas they cover. Publications can be in one or more content areas.

Figure 11
Content Area Chart of All DART Publications Found in the Web of Science
(# of publications=101)

As expected, the majority (62%) of DART publications were categorized in the Computer Science content area. Other content areas these publications covered include Mathematics (23%); Engineering (22%); Mathematical Computational Biology (20%); Communication (18%) and Genetics Heredity (18%).

Figure 12 shows year of publication and number of citations for the 101 publications.

Figure 12
Citation Analysis of DART Publications Found in the Web of Science
(# of publications=101)



These 101 publications were cited in 412 articles for an average of 4.3 times per publication. The H-Index of the DART faculty was 9. The H-index or Hirsch index is a metric for evaluating the cumulative impact of an author or group of authors scholarly output and performance. It measures quantity with quality by comparing publications to citations. A group of researchers with an h-index of 9 have published nine papers that have been cited at least nine times by other scholars.

Figure 13 lists those publications that had an average of more than five citations per year.

Figure 13
Publications with an Average of Five or More Citations Per Year

Publication Title	EPSCoR Authors	Year Published	Journal	Average # of Citations Per Year	Total Citations
Machine Learning Methods in Drug Discover	David Ussery, Xiuzhen Huang	2020	Molecules	18	72
Advances in Data Preprocessing for Biomedical Data Fusion: An Overview of the Methods, Challenges and Prospects	Emre Celebi	2021	Information Fusion	16.7	50
Image synthesis with adversarial networks: A comprehensive survey and case studies	Emre Celebi	2021	Information Fusion	14.3	43
Sulforaphane prevents age-associated cardiac and muscular dysfunction through Nrf2 signaling	Se-Ran Jun	2020	Aging Cell	10.8	43
Deep reinforcement learning in medical imaging: A literature review	Thi Hoang Ngan Le, Khoa Luu	2021	Medical Image Analysis	10	30
Deep reinforcement learning in computer vision: a comprehensive survey	Thi Hoang Ngan Le, Khoa Luu	2022	Artificial Intelligence Review	7	21
Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology	Emre Celebi	2022	Jama Dermatology	6	18

Patents

There have been six reported patent disclosures filed since the start of the project.

Figure 14
Disclosures and Patents

Year	Title	Researchers	Component
3	Multiple Camera Multiple 3D Object tracking on the Move for Autonomous Vehicles	Khoa Luu	Learning & Prediction
3	Self-supervised Domain Adaptation in Poultry Counting in the Wild	Khoa Lee	Learning & Prediction
3	Smart Insect Control Device via Artificial Intelligence in Realtime Environment	Khoa Luu, Tanh-Dat Truong	Learning & Prediction
3	Sensor-based Smart Insect Monitoring System in the Wild	Khoa Luu, Tanh-Dat Truong	Learning & Prediction
3	Artificial Intelligence and Vision-Based Broiler Body Weight Measurement System and Process	Thi Hoang Ngan Le, Khoa Vo	Learning & Prediction
3	Chicken Processing Plant with Automated Computer Vision	Thi Hoang Ngan Le, Khoa Vo	Learning & Prediction

External Engagement (Scientific Literacy, Public Presentations, Policymakers, Education)

Increasing the scientific literacy and understanding of scientific research at all levels of society is important for increasing the diversity of the scientific workforce. This includes the general public, undergraduates, graduate students, junior faculty, K-12 teachers and others.

Figure 15 shows the number of people who have been engaged by type of audience with these outreach efforts during the project.

Figure 15
Number of People Reached through
External Engagement Efforts for All Program Years

Type of Institution	**Academic Research Institutions**		**Primarily Undergraduate Institutions**		**Minority Serving Institutions**		**K-12 Institutions**				**Total**
Type of Audience	Faculty	Students	Faculty	Students	Faculty	Students	Teachers	Students Reached Directly	Students Reached via Teacher Training	Other	
Male	44	38	18	21	10	13	13	540	0	10	707
Female	26	29	27	12	4	13	61	575	0	10	757
Underrepresented Minority	3	3	2	8	3	23	10	399	0	0	451
Total	71	67	45	33	14	26	114	1,311	0	181	1,862
Percent	4%	4%	2%	2%	1%	1%	6%	70%	0%	10%	

Overall, an estimated 1,862 people have been involved in one or more external engagement activities supported by DART during the project. More than two-thirds (70%) were K-12 students reached directly. Two-thirds (44%) of the students reached directly were female and 30% were an underrepresented minority in STEM.

Findings by Component

1. **Coordinated Data Science Cyberinfrastructure – The Arkansas Research Platform (ARP)**

Proposed: *"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 Arkansas Research Platform ARP, and integrate it with existing high-performance computing 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 stream data to visualization environments both proximate and distant from the computing resources to aid in analysis and meta-analysis of experiments. ARP will be managed as a unique multi-institutional resource."*

Findings:

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	0	2	0	2
Joint Publications*	0	3	0	3
Presentations	3	0	2	5
Joint Presentations*	1	0	0	3
Workshops	0	6	1	7
Proposals # ($)	4 ($1.1M)	2 ($3.4M)	1 ($100k)	7 ($4.7M)
Awards # ($)	1 ($100k)	2 ($784k)	0	3 ($884k)
Joint Proposals* # ($)	1 ($800k)	2 ($5M)	0	3 ($5.8M)
Joint Awards* # ($)	1 ($800k)	1 ($95k)	0	2 ($895k)

*Collaborative proposals/awards/publications with other DART components

Establish the Arkansas Research Platform as a shared data science resource across the jurisdiction.

Year 1

-established CI working group and held monthly meeting
-developed CI plans for UAF and UAMS
-issued purchase orders for:
  -20 nodes dual AMD 7543, 1024GB, NVMe local drive, single PCI 40GB A100GPU
  -4 nodes dual AMD 7543, 1024 GB, NVMe local drive, four SXM 40 GB A100 GPU
  -100 Gb Inifiniband connection and 10Gb Ethernet connection
  -3 Enclosed cooled racks
-provided access to ARP through Science DMZ at UAF
-implemented Gitlab with dedicated server behind Science DMZ at UAF
-established Globus Basic server at UAF with endpoints at storage arrays at UAF and UAMS
shared via a OneDrive to participants

Year 2

-all nodes became operational with network changed to the UA ScienceDMZ available to all DART researchers
-increase of 73% in ARP resources; now have 1,374 teraflops available
-established secure method of providing access to ARP resources at UAF and UAMS
-participants gain access to interactive sessions on nodes via Open OnDemand portals, Pinnacle and Grace clusters
-access to storage arrays at UAF and UAMS available through Globus as endpoints
-UAF ITS moved to commercial cloud hosted GitLab repository
-installed Git on Pinnacle and Grace - researchers can clone or copy from DART GitLab via SSH
-re-budgeted Globus licenses in year 1 and 2 to address federated identity management solutions
-created prototype secure enclaves at UAMS and experimentation using container-based approach with Kubernetes orchestration
-developed System Security Plans to host HIPAA and Controlled Unclassified information at UAF
-created and filled postdoctoral fellow position
-publish UAF CI design and configuration document on website
-host 5 software carpentry workshops and train 2 carpentry instructors
-host 2 online ARP training sessions

Year 3
-working prototype using UA Systems Bridge to automatically authenticate users
-host 5 carpentry workshops and 2 online ARP training sessions
-procedures for managing CUI have been implemented
-created containerized Hadoop-based testbed for DC
-installed 100gb switch for UAMS
-draft user guidelines and policies

Visualization for complex data in diverse data-analytics application domains.
Year 1
-none reported

Year 2
-conducted systematic literature review on advanced visualization and immersive analytics
-results internally reviewed and report is ready for publication on DART website
-incorporated data-driven modeling and photogrammetry-based visualization into virtual-reality experiences for undergraduate geoscience education
-published results of a user evaluation on different design choices of virtual field trips in Journal of Educational Computing Research
-collaborated with NY State U at Albany to examine effects of augmented-reality display on windshield of self-driving vehicles for drivers' spatial awareness
-designed a virtual collaborative space in augmented and virtual reality to improve coordination between data science educators and industry partners in Arkansas
-developed an immersive workbench to visualize and analyze environmental data collected from rural Borneo Highland for use in ecological research and education
-host 1 online advanced visualization workshop

Year 3
-activities delayed due to loss of personnel

Conclusions/Recommendations
The Coordinated Data Science Infrastructure component has been successful at acquiring and installing hardware, but less so in providing access to that hardware for users. The team has purchased and installed: 20 nodes dual AMD 7543, 1024GB, NVMe local drive, single PCI 40GB A100GP; 4 nodes dual AMD 7543, 1024 GB, NVMe local drive, four SXM 40 GB A100 GPU and 100gb switch for UAMS. The visualization for complex data in diverse data-analytics application domains has been delayed by loss of personnel.

The issue of federated identity management across participating institutions is being addressed by using UA Systems Bridge to automatically authenticate users. The team still needs to fully automate this process without compromising security. Workshops on using an interactive shell to access Pinnacle are helping to increase the use of these computer resources by faculty and graduate students. A one-on-one connection between IT professionals and researchers seems the best approach for increasing HPC usage.

The evaluator has some concern that the goal on visualization for complex data will be able to be completed. The loss of personnel and the additional work needed to implement federated identity management may impact its completion.

2. Data Life Cycle and Curation (DC)

Proposed: *"Our research will aim to increase the level of automation in data curation and governance. We will explore a closed-loop data analytics approach, emphasizing need- and prediction-based data collection and transmission as well as the feedback role of current decisions on future data collection. Our research will encompass data governance, data architecture, data integration, and data quality, using ARP to implement and test tools to implement these ideas."*

Findings:

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	4	19	11	34
Joint Publications*	0	2	0	2
Presentations	13	15	16	44
Joint Presentations*	1	0	0	1
Workshops	3	0	0	3
Proposals # ($)	8 ($9.7M)	12 ($3.9M)	2 ($919k)	22 ($14.5M)
Awards # ($)	1 ($436k)	5 ($1.1M)	6 ($4.8M)	12 ($6.4M)
Joint Proposals* # ($)	2 ($1.5M)	2 ($5.1M)	0	4 ($6.6M)
Joint Awards* # ($)	0	2 ($1.6M)	0	2($1.6M)

*Collaborative proposals/awards/publications with other DART components

DC1: Automate heterogeneous data curation

Year 1

-implemented cluster entropy metrics in the Data Washing Machine (DWM) proof-of-concept

Year 2

-progress on Python POC for DWM code in 10th release, version 2.21; improvements to code improve precision and recall values linking 19 base data sets
-implemented robotic process to perform grid search across DWM parameters to find setting to produce best precision and recall clustering results
-research developing new unsupervised data quality assessment for data redundancy based on "cluster quality metric"
-use NLP models, BERT and RoBERTa, in a Zero-Shot Learning context, to project records into latent vector spaces and measure distances or similarities
-computed metrics used to cluster references using ML algorithms
-created visual representations of topological data analysis of selected test data sets
-converted textual data converted to vector spaces and simplified and segmented using topology-based tools
-continued implementation of collaborative/need-based data collection mechanism in disaster relief decision making
-developed basic data query and collection schemes for image data collection from Google Street View
-conducted initial experiments on damage assessment on a sample image dataset previously collected from social media
-developed a method for automatic data collection from Twitter's open API
-developed basic data query and collection scheme for tweets on a disaster event
-investigated the genetic variations underlying drug response, studied single-cell RNA sequencing (rna-seq), data of chronic-phase chronic myeloid leukemia stem cells
-developed tyrosine kinase inhibitors (TKI) to target the BCR-ABL oncoprotein
  -inhibiting its abnormal kinase activity
  -significantly improved CML patient outcomes
  -cells with distinct responses to TKI, good vs poor, were clustered together in both BCR-ABI positive and negative cells
  -showed putative transcription factors of these expressed genes revealed by single-cell regulatory network inference and clustering

Year 3

-created knowledge graphs based on data collection objectives
-developed a need-based sampling methodology to train surrogate models for estimating terminal network reliability
-extended this framework to other supervised learning contexts
-completed development of new data imputation method for missing values recovery
-submitted manuscript for publication
-employed linear & nonlinear dimensionality reduction methods to visualize & reveal distinct clusters in high dimensional datasets

DC2: Explore secure and private distributed data management

Year 1

-built and compared classification models using classical and deep learning algorithms for Alzheimer's disease prediction and biomarker identification
-created and maintained GitHub repository for the DWM including proof-of-concept code
-created a separate ground-truth file to compare results of any algorithm applied to data
-setup framework and tested to extract data from public repository of all Federal contracts and awards
-conducted preliminary study for unlabeled paired samples using the Min-Max ratio test

-documented DWM POC process global (file-level) unsupervised data cleansing methods shared with team members on DWM working paper/publication and Python code on BitBucket.org
- used artificial intelligence model trained for NLP tasks to find multidimensional vector embeddings for text record in data sample
- evaluated the similarity of two text records using a Natural Language Inference tool
-created an implementation of the transitive closure Java routines in Python
-implemented a global (file-level) data cleaning routine comparing high-low frequency tokens
-developed novel algorithm, SCAN to detect clusters, hubs, and outliers in networks
-SCAN clusters vertices based on a structural similarity measure
-published SCAN in ACM SIGKDD'07 received over 821 citations (ref. Google Scholar)
-results shows a superior performance when using novel machine learning and AI models for algorithms to automate data cleaning
-DWM Refactor codes available on BitBucket.org. DWM POC is being refactored as a Python Program.

Year 2
-automation of data cleansing expanded to include data corrections based on record-to-record comparisons with blocks and clusters
-developed 7 techniques for record-to-record correction, some can impute missing values and correct incorrectly split or joined tokens
-developed novel framework for scalable Entity Resolution using NLM, Locality Sensitive Hashing and Machine Learning
-preliminary experiment result shows promise, achieving accuracy over 95% with a nearly linear runtime
-developed computational framework integrating multi-layer genomics data to identify transcriptome and pathway dysregulations in autism spectrum disorder
-inferred regulatory networks differentially expressed in the disease sample as compared to control samples
-provides a way to reveal master regulators which position at top of regulatory hierarchies and control the transcriptional activities of many downstream genes
-established regulatory cascades approach offering a framework for revealing new disease-related genes
-can be applied and extended to study other tissues and diseases
-investigated the expression alterations of survival-related genes in various immune cell types when combining breast cancer bulk and single-cell RNA sequencing data
-helps to better understand the interactions of tumor and immune systems
-provides novel molecular prognostic markers for survival prediction in breast cancer patients
-developed method can be applied to study other types of cancer
-developed variety of genome visualization tools ("R-BioTools") used in courses and workshops living in a GitHub directory at UAF
-automated quality scores for biological sequence data applied to viral and bacterial genomes
-SARS-CoV-2 published February 2022, using automated genome quality scores in helping to cluster genome species
-automated pipeline tested for Enterococcus faecium, Salmonella enterica, and Klebsiella pneumoniae
-developed a convolutional neural network-based model to recover missing values for scRNA-seq data
-probability of dropout computed using gamma-normal expectation maximum algorithm
-model demonstrated robust performance, achieving comparable or better results compared to other imputation methods
-identified novel pattern related to daptomycin resistance through bid data analysis of genomes
-novel pattern suggests a new paradigm of daptomycin resistance dissemination
- established a computational workflow of several combined machine learning approaches to identify biomarkers for:
-prostate cancer using metabolomics data,
-chemotherapy-induced cardiotoxicity among breast cancer patients using metabolomics data
-investigating potential new antibiotic resistance genetic markers using known markers in Enterococcus faecium using machine learning approach and population structure of species

Year 3
-validated proof of concept for positive data control system
-expanding focus to include data governance and other aspects of data life cycle in response to EAB and RSV reports

DC3: Harmonize multi-organizational and siloed data
Year 1
-downloaded more than 300k bacterial and archaeal genomes from the NCBI and complete set of genomes from Integrated Microbial Genomes and Microbiomes project
-built a structured, organized genome database stored on the high-performance computer at UAMS
-performed quality score analysis of the bacterial genomes in GenBank
-used a standardized pipeline for finding genes for each genome
-developed and published a program, ProdMX, to speed up genome comparison more than a million-fold compared to traditional alignment methods

-used heat maps to visualize the comparisons of more than hundred-thousand E. coli genomes
-created a matrix of distances displayed as a heat map utilizing Mash and in-house Python script
-developed methods and tools using R-BioTools for visualization of pan- and core-genomes
-released python pipeline for pan- and core-genome-based functional profile for metagenomics samples from microbial communities
-tested three commonly used sequenced-based methods for predicting an organism's taxonomy
-utilized bacterial genomes to build a machine learning approach for meta-proteomics analysis
-published a paper giving an overview of multi-omics approaches in the journal "Molecular Omics"
-developed a multi-omics data integration pipeline consisting of DNA methylation, mRNA, protein, phosphopeptides and histone post-translational modifications to understand the regulation of triple negative breast cancer subtypes MDA-MB-231 (BRCA1wt) and HCC1937 (BRCA15382insC) cell lines

Year 2
-developed integrated database for proteomics & genomics, including annotations
-published a R-Bio Tools paper for visualizing genomes
-developed ML models for known toxins

Year 3
-developed clean ESKAPE genome database using GRUMP
-four publications, including in BioRxiv
-explored using XGBoost for antibiotic resistance prediction
-developed application comparable to existing ML approaches
-developed and published an approach to study drug resistance in CML

Conclusions and Recommendations
The Data Life Cycle and Curation component is meeting most of its objectives. Faculty have a prototype version of the washing machine; automated data cleansing to include data corrections based on record-to-record comparisons with blocks and clusters; developed computational framework integrating multi-layer genomics data to identify transcriptome and pathway dysregulations in autism spectrum disorder; investigated the expression alterations of survival-related genes in various immune cell types when combining breast cancer bulk and single-cell RNA sequencing data; established a computational workflow of several combined machine learning approaches to identify biomarkers for both prostate cancer using metabolomics data and chemotherapy-induced cardiotoxicity among breast; downloaded more than 300k bacterial and archaeal genomes from the NCBI and complete set of genomes from Integrated Microbial Genomes and Microbiomes project; developed and published a program, ProdMX, to speed up genome comparison more than a million-fold compared to traditional alignment methods; and published a paper giving an overview of multi-omics approaches in the journal, "Molecular Omics" among other accomplishments.

It is commendable that the team is incorporating suggestions made by the EAB and RSV into its work. It is not clear that other components in DART or others in the data science community are using any of the algorithms or programs/processes developed by this research group. Recommend that this group do more to advance the use of their algorithms, perhaps holding a workshop for DART faculty and graduate students and tracking the use of DC algorithms in the broader data science community. The DWM programs may be of interest to those college students in some of the new data science programs being developed around the state.

3. Social Awareness (SA)

Proposed: *"In this research area, we focus on developing cutting-edge, socially aware data analytics to address social concerns and meet laws and regulations in national-priority applications and better enable big data analytics to promote social good and prevent social harm. In particular, we will address the following critical challenges. How do we: 1. achieve meaningful and rigorous privacy protection when mining private data or collecting sensitive data from individuals? 2. ensure non-discrimination, due process, and understandability in decision-making? 3. achieve safe adoption, and robustness of machine learning and big data analytics techniques, especially in adversarial settings? 4. incorporate social awareness in domain- or application-specific projects? Our research goals are to develop novel techniques to provide privacy preservation, fairness, safety,and robustness to a variety of data analytics and learning algorithms including automated data curation, social media and network analysis, and deep learning, and ensure the adoption of the developed techniques meet regulations, laws and user expectations."*

Findings:

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	2	11	14	27
Joint Publications*	0	0	0	0
Presentations	7	2	4	13
Joint Presentations*	0	0	0	0
Workshops	0	0	1	1
Proposals # ($)	4 ($3.2M)	6 ($3.9M)	2 ($1.2M)	12 ($8.3M)
Awards # ($)	0	3 ($725k)	2 ($1.2M)	5 ($1.9M)
Joint Proposals* # ($)	2 ($4.7M)	1 ($750k)	0	3 ($5.4M)
Joint Awards* # ($)	0	2 ($1.45k)	0	2 (1.45k)

*Collaborative proposals/awards/publication with other DART components

SA1: Privacy-Preserving and Attack Resilient Deep Learning

Year 1

-conducted a survey of existing attacks on deep learning models broadly categorized into evasion, poisoning and model stealing
-researched representative algorithms, using threat models from four aspects: adversarial falsification, adversary's knowledge, adversarial specificity, and attack frequency
-conducted theoretical studies of the potential risks of deep learning models and privacy preserving mechanisms for deep learning
-completed literature review on definition of personal identification information (PII) from different perspective and privacy issues
-proposed frameworks to assess sensitivity of information in different contexts
-developing metric to consider sensitivity level of each PII attribute and combined sensitivity of a given set of leaked PII attributes

Year 2

-designed framework to generate poisoning samples to attack model accuracy/algorithmic fairness of fair machine learning models
-developed 3 online attacking methods: adversarial sampling, adversarial labeling, and adversarial feature modification
  -all effectively and efficiently produced poisoning samples of training data to reduce the test accuracy
  -can flexibly adjust the attack's focus and accurately quantify the impact of each candidate point to accuracy loss and fairness violation, producing effective poisoning samples
  -conducted experiments on two real datasets demonstrating the effectiveness and efficiency of our attacking framework
-studied privacy preserving mechanisms used for deep learning algorithms
  -studied the inequality in utility loss due to differential privacy
  -compared changes in prediction accuracy w.r.t. each group between private and non-private model
  -analyzed cost of privacy w.r.t. each group, explain how group sample size and other factors relate to privacy impact
  -examined the privacy, resilience, utility tradeoff of deep learning models
  -developed threat-and-privacy-aware deep learning models
  -developed a modified DPSGD algorithm called DPSGD-F to achieve and equal costs of differential privacy and good utility
  -conducted experiments on real world datasets
  -results showed the effectiveness of DPSGD-F algorithm on achieving equal costs of differential privacy with satisfactory utility
-developed a novel adversarial adaptive defense (AAD) framework based on adaptive training
  - showing trained models adapt at test time to new adversarial attack
  -framework improved structures of training data into groups and each group represents one attack scenario
  -learns a context vector from features of each batch during training
  -incorporates the learned context vector into both prediction and detection models
  -conducted evaluations with popular adversarial attacks and defense strategies on two real world datasets under different attack settings with results showing AAD achieves high prediction and detection accuracy and significantly outperforms baseline
-developed robust framework under distribution shift that adopts reweighing estimation approach for bias correction and minimax robust estimation approach for achieving robustness on prediction accuracy

Year 3

-explored tradeoffs among privacy, resilience, and utility in deep learning models
-developed novel heterogenous randomized response mechanism to protect nodes' features and edges against privacy inference attacks under differential privacy
-achieved significantly better randomization probabilities and tighter error bounds with this method than from existing approaches
-investigated the application of randomization mechanisms in high-dimensional feature settings
-conducted extensive theoretical and empirical analysis using benchmark datasets that showed our approach significantly outperforms various baselines in terms of privacy protection for both nodes' features and edges

SA2: Socially Aware Crowdsourcing

Year 1

-investigated interval-valued labels to enable a worker to specify both type-1 and type-2 uncertainties in his/her label without information loss

-developed algorithms to aggregate labels as an inference with a preferred probability of matching above 50% computationally

-developed strategies to pre-process collected interval-valued labels into two categories, data cleaning and normalization

-extended traditional statistic and probabilistic concepts for point-valued datasets to interval-valued including mean, variance, standard deviation, and probability density function for interval-valued labels.

-investigated two learning algorithms on deriving inferences from interval-valued labels using majority voting and preferred matching probability

-performed computational experiments to test the effectiveness of applying interval valued labels in managing uncertainty in crowdsourcing

-experiments successfully verified theoretical and algorithmic results

-paper accepted by 2021 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS 2021

Year 2

-applied interval-valued labels (IVL) instead of binary-valued

-worker may use a subinterval within (0,1) to annotate an instance even when uncertain

-developed two algorithms (i.e., internal-valued majority voting (IMV) and preferred matching probability (IPMP) to derive inferences from interval-valued labels

-computational experiments evidence the proposed interval-valued scheme enables specification of uncertainties during input time.

-IMV and IPMP algorithms computationally derive an inference with above 50% probability of matching the ground truth

-uncertainty index defined in work quantitatively measures overall uncertainty of collected IVLs

-produce better quality inferences with IVLs than without

Year 3

-activities delayed due to loss of personnel

SA3: User-centric Data Sharing in Cyberspaces

Year 2

-documented and disseminated research on identification techniques for non-structured data

Year 2

-continued exploration and development of techniques for identifying context aware sensitive information from unstructured data

-researching multimodal deep learning techniques for detecting and removing sensitive information, discriminating and stigmatizing information from unstructured data

Year 3

-activities delayed due to loss of personnel

SA4: Deep Learning for Preventing Cross-Media Discrimination

Year 1

-conducted theoretical investigation on CNN deep learning models to identify discriminatory objects from social images

-conducted research on adopting long, short-term memory network to model text (captions, tags and discussion of social images)

-conducted research and empirical studies on multimodal hate speech detection

-detecting coded words in hate speech detection

-developed coded hate speech detection framework, CODE, to judge coded words used in the coded meaning

Year 2

-developed a deep learning-based coded hate speech detection framework called CODE

-CODE findings published and presented to research community

-proposal submitted on CODE findings

-conducted empirical analysis on multimodal hate speech detection models

-evaluated performance of Facebook Hateful Meme Challenge baseline models on 3 MMHS150K datasets, image and text inputs

-trained models using different baseline approaches, unimodal training, multimodal training with unimodal pretraining and multimodal pretraining

-evaluated metrics, accuracy and the Area under the ROC Curve (AUROC)

-evaluation shows current multimodal training does not significantly outperform unimodal training

Year 3
-activities delayed due to loss of personnel

SA5: Marketing Strategy Design with Fairness
Year 1
- performed exploratory study on fairness-aware design decision-making
  -trained Logistic regression and CatBoost classifiers on the pre-processed dataset to predict individuals 'income in test data using 10-fold cross-validation approach
  -conducted Disparate Impact (DI) analysis, observing gender attribute and ethnicity attribute ducted fairness tested based on the calibration scores using probability score to determine gender attribute discrimination

Year 2
-conducted link prediction in identity network based on social network, intra-layer and inter-layer link information
  -can predict number of nodes affected in the entire social network
  -conducted comparison with theoretical approaches, independent cascades and liner threshold
  -quantified unfairness and analyzed its impact in the context of data-driven engineering design using the Adult Income dataset
  -introduced standard definitions and statistical measure of fairness to the engineering design research
  -used outcomes from 2 supervised machine learning models, Logistic Regression and CatBoost classifiers
  -conducted disparate impact and fair-test analyses to quantify unfairness present in the data and outcomes
  -findings published and presented to research community

Year 3
-activities delayed due to loss of personnel

SA6: Privacy-Preserving Analytics in Health and Genomics
Year 1
-conducted survey of existing frontier work and investigation of mathematical models related to privacy-preserving analysis
-algorithms are built based on models related to computational phenotyping, mathematical optimization and statistics models

Year 2
-documented and disseminated findings of literature research of privacy-preserving data analytics algorithms and software
-initiated investigation on mathematical optimization models
-findings published and presented to research community

Year 3
-activities delayed due to loss of personnel

SA7: Cryptography-Assisted Secure and Privacy-Preserving Learning
Year 1
-conducted survey of existing work using cryptography for privacy protected in federated learning
  -analyzed each work using studied machine learning model/algorithm/method, type of dataset partition, cryptography method and whether differential privacy is provided
  -designed a new cryptography-based scheme for differentially privacy federated learning
  -established two goals for the new cryptography-based scheme
  -reduce communication cost in training process and improve the learning accuracy while providing differential privacy
  -experimental results show scheme performs better than existing work in convergence rate and learning accuracy

Year 2
-designed a cryptography-based solution
-developing a privacy-preserving face recognition-based access control system

Year 3
-activities delayed due to loss of personnel

Conclusions and Recommendations
The Social Awareness component activities and progress have been impacted by the loss of key personnel. Prior to Year 3 this component had researched representative algorithms, using threat models from four aspects: adversarial falsification, adversary's knowledge, adversarial specificity, and attack frequency; developing metric to consider sensitivity level of each PII attribute and combined sensitivity of a given set of leaked PII attributes; developed a novel adversarial adaptive defense (AAD) framework based on adaptive training; investigated interval-valued labels to enable a worker to specify both type-1 and type-2 uncertainties in his/her label without information loss; developed coded hate speech detection framework, CODE, to judge coded words used in

the coded meaning; performed exploratory study on fairness-aware design decision-making; conducted link prediction in identity network based on social network, intra-layer and inter-layer link information; documented and disseminated findings of literature research of privacy-preserving data analytics algorithms and software; conducted survey of existing work using cryptography for privacy protected in federated learning and designed a cryptography-based solution.

As recommended by the RSV this research group is revising their milestones and exploring more crossover with the Social Media team. The research team will be submitting a revised strategic plan for Years 4 and 5 in the coming weeks.

4. Social Media and Networks (SM)

Proposed: *"Our research will address these challenges in close collaboration with Wu and Sheng (SA) by exploring innovative methods and techniques of mining argumentation data in social networks and analyzing its characteristics, such as polarization, opinion diversity, participant influence, opinion community, and opinion prediction; transformative multilayered network analytic method to analyze deviant behaviors in social media networks by modeling multi-source, supra-dyadic relations, and shared affiliations among deviant groups; and innovative algorithms for logistics planning in disaster response using social media platforms. The models and insights generated from the proposed research will enhance our ability to both capitalize on the potential of social media as a force of good while mitigating its use as a weapon."*

Findings*:*

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	0	28	9	27
Joint Publications*	0	0	0	0
Presentations	12	10	12	34
Joint Presentations*	0	0	0	0
Workshops	0	0	3	3
Proposals # ($)	9 ($30.1M)	2 ($1.2M)	3 ($1.1M)	14 ($32.5M)
Awards # ($)	5 ($3M)	2 ($8.7M)	3 (1.3M)	11 ($12.9M)
Joint Proposals* # ($)	0	0	0	0
Joint Awards* # ($)	0	0	0	0

*Collaborative proposals/awards/publications with other DART components

SM1: Mining cyber argumentation data for collective opinions and their evolution

Year 1

-determined key featured for platform

-prepared software design guideline document for platform development

-developed and designed new social issue generator for social network data collection processing

-received IRB approval for social issue generators for Fall 2020 Data Collection

-determined baseline individual user and Social Network measure for cyber-discourse platform

- used Intelligent Cyber Argumentation System (ICAS) platform

-collected social network data to identify key variables and mine data to populate statistical compatible datasets (SPSS) from cohort of students in 2018-2020 General Sociology classes

Year 2

-mine thousands of lines of sentences from ICAS platform

-developed novel algorithm

-testing algorithm for evaluation

Year 3

-activities delayed due to loss of personnel

SM2: Socio-computational models for safer social media

Year 1

-identified social media platforms used in different cyber influence campaigns and different contexts and geographical regions

-identified characteristics and features of social media platforms

-conducted multi-taxonomy characterization of social media data

-identified data sources

-developed and published data collection methodology
-created database schema to accommodate new fields with changes in data sources or characteristics
-identified data acquisition methods to include API access and web scraping
-procured an academic data collection license (Twitter)
-submitted data access proposal and review which was accepted

-developed, tested, and deployed data collection framework with real-time dashboard to monitor progress with alerting capabilities
-identified key characteristics of classification, perceptive quality, and scalability of multimedia data on social platforms
-identified one of the three major learning objectives
-identified damage assessment and verification based on image and video data in disaster response as one of the three key applications
-published work on identifying cyber campaigns and features, and established data acquisition procedures in collaboration with practitioners and policy makers within and outside Arkansas

Year 2

-revised taxonomy to characterize OIE based on social media platforms; studied cyber campaigns and characteristics of platforms and involved information actors
-developed socio-computational model
-identified focal structures, leverages theory of social network analysis/collective action and uses operations research framework
-evaluated with data collected on YouTube conspiracy theory spreaders and Twitter misinformation networks
-datasets correspond to application areas, COVID-19, smart city infrastructure security protests, social movements
-devised indexing methods for image, video and associated meta and text data
-identified key applications of "smart" use of multimedia information sources
-use of information quality aspects as part of future smart data-based applications
-focus on reliability dimension

Year 3

-development and publication of a taxonomy to characterize social media
-collected and published various cyber campaign datasets
-developed and published model to identify contextual focal structures fundamental to coordinating cyber campaigns
-refined model with multiplex representation of individuals and contexts to offer explainability/interpretability of findings

SM3: Auto-annotation of multimedia data

Year 1

-completed two disaster response routing problem variants using Milburn's existing qualitative interview data
-completed reviews of the CTP and OP academic literature

Year 2

-began devising indexing methods for image, video and associated meta and text data
-key applications of "smart" use of multimedia information sources have been identified with a focus on the reliability dimension

Year 3

-identified multimedia data characteristics towards target applications
-integrated learning objectives from SM4 into disaster management scenarios

SM4: Informing disaster response with social media

Year 1

-working to identify content types on social platforms to describe transportation infrastructure status after disruptions due to a disaster
-documented the workflow surrounding how to manually transform data from individual content elements posted to social platforms into transportation infrastructure status information

Year 2

-selected Hurricane Harvey for disaster scenario and will be primary focus for initial scenarios
-represented a large-scale and geographically widespread event
-began assembling disaster image dataset from online disaster image repositories
-downloaded tweets and images from Twitter API
-uses content-based techniques to verify or complete the metadata associated with visual data, location, orientation, timestamps and identification of major landmarks
-gathered initial sample data set for Harvey and performed an analysis and evaluation of techniques
-selected 2 routing problems with application to disaster response

-literature review on disaster logistics problems and models
-deployed PostGIS database and hosted by CAST

Year 3
-multimedia data from two major SM sources have been identified, collected and analyzed
-one publication on water depth estimation and one on speed estimation based on social media data
-building methodology to index transportation infrastructure for a given geographical area
-obtained data from Hurricane Harvey as proof of concept
-evaluated multiple approached for using images from SM for identification of flood afflicted areas
-routing algorithms were developed, validated and tested for single and multiple agent and destination variations of an uncertain and disrupted network

Conclusions and Recommendations
The Social Media and Networks component has lost some key personnel and is working to update the strategic plan and work more closely with Social Awareness as recommended by the RSV. This component has determined key features and prepared the software design document for the cyber social network platform; developed, tested, and deployed data collection framework with real-time dashboard to monitor progress with alerting capabilities; revised taxonomy to characterize OIE based on social media platforms; studied cyber campaigns and characteristics of platforms and involved information actors and obtained data from Hurricane Harvey to represented a large-scale and geographically widespread disaster scenario.

Component researchers have produced 27 publications, 34 presentations and been awarded over $11 million dollars to advance their research. These publications include a taxonomy to characterize social media; various cyber campaign datasets; a model to identify contextual focal structures fundamental to coordinating cyber campaigns; one publication on water depth estimation and one on speed estimation based on social media data and routing algorithms used for single and multiple agent and destination variations of an uncertain and disrupted network. The progress in developing models and algorithms that can use Social Media postings to accurately inform disaster response is especially noteworthy.

The progress in SM1: Mining cyber argumentation data for collective opinions and their evolution and SM3: Auto-annotation of multimedia data are harder to gauge. The team reports the loss of key personnel and "human resource issues for graduate students" as partly to blame for lack of progress in these two areas. Perhaps revising the milestones and/or combining efforts with the Social Awareness team may be appropriate.

5. Learning and Prediction (LP)

Proposed: "*Research in this topical area will focus on various techniques in prediction interpretation for large-scale, deep learning using multi-source integrated data sets. In particular, we will focus on applying statistical learning techniques alongside more advanced deep learning techniques to address three major challenges. 1)Violation of fundamental statistics principles. 2)Mode specification and interpretation. 3)Computing in big data environments. We will investigate these challenges surrounding high-dimensional, dynamic and unstructured data sets and explore solutions in the domains of genomics, transaction scenarios in eCommerce, and supply chain logistics."*

Findings:

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	2	26	11	39
Joint Publications*	0	1	0	1
Presentations	6	13	20	39
Joint Presentations*	0	1	0	1
Workshops	1	1	2	4
Proposals # ($)	3 ($2.8M)	10 ($3.8M)	5 ($3.5)	18 ($10.1M)
Awards # ($)	1 ($1.2l)	1 ($50k)	3 ($195k)	5 ($246k)
Joint Proposals* # ($)	3 ($2.3M)	2 ($300k)	0	5 ($2.6M)
Joint Awards* # ($)	1 ($800k)	3 ($1.7M)	0	4 ($2.5M)

*Collaborative proposals/awards/publications with other DART components

LP1: Statistical Learning – Random Forests for Recurrent Event Analytics
Year 2
-transitioned MTPP and LTSM to a convolutional neural network (CNN) approach
-implemented completely in Keras
-established pipeline to allow maintenance based tabular data set first tested on model
-published work on CNN for image-based data

-work serving as foundation for shift in methodology from MTTP/LSTM integration

Year 3
-established a random forest-based model for recurrent event data implemented in R
-paper submitted for publication

LP2: Statistical Learning – Marked Temporal Point Process Enhancements via Long Short-Term Memory Networks

Year 1
-replicated one of the existing methods in the literature for intelligent food-borne disease investigation based on event data
-explored the RF-SRC method and use of NHPP to model the recurrent event
-literature review identified gaps in existing methods
-curated dataset of sensor data, system attributes, and failure/repair data of 8332 oil/gas wells installed 2007-2021

Year 2
-collected civil infrastructure datasets publicly available
-replaced curated healthcare loT datasets with curated maintenance-based tabular data-curated and available
-storing datasets on shared repository via the cyberinfrastructure team

Year 3
-benchmarking our approach against other RNN architectures on failures predictions

LP3: Deep Learning: Novel Approaches
Year 1
-developed library of classifiers on natural images from simple linear classifiers
-SVMS and linear autoencoders to complex deep learning approaches
-AlexNet, ResNet, and GoogleNet, and standard machine learning classifiers
-investigated efficacy of group structure on generalized neural network (GNN) architecture with smallest finite simple nonabelian group A5 action of random and clustered synthetic small size data
-applied techniques to the color channel data of three-dimensional fundamental topological structures
-resulting data studied for any significance by exploratory data analysis and statistical inferential methods
-development of generalized model of reward function in DRL addressing issues with sparse and dense feedback
-developed multiple low-cost deep learning methods
-Teacher-Student Distillation Deep Learning, Distilled ShuffleNet, Self-Knowledge Distillation Algorithms

Year 2
-investigating the injectivity issues in persistent homology
-2 distinct shapes have identical topological representation making them unusable as discriminating features in neural network
-studied a use-case with "Montezuma's Revenge, Atari game for a DRL agent
-using support of human intuitions (in from of heuristics), agent was able to perform exploration very quickly
-submitted paper on work to a conference

Year 3
-personnel changes will require strategic plan revisions
-developed a novel self-supervised domain adaptation method in crowd counting
-published two novel tools-Fairness Domain Adaptation (FREDOM) and Approach to Semantic Scene Understanding and self-supervised Spatiotemporal Transformers (SPARTAN) available om GitHub
-published research paper in IEEE ACCESS detailing novel features for classifying malware using machine learning models
-published an image preprocessing model at the Intl. Conference on Image processing Computer Vision & Pattern Recognition
-submitted paper to Future Technologies Conference on comparative analysis of object localization using topological data
-multiple causal inference models have been studied and are being investigated on their use in DRL

LP4: Deep Learning: Efficiency and Specification
Year 1
-analyzed current issues of distillation methods for computer vision dataset
-face recognition, action recognition, and medical imaging
-processing a private insurance claims dataset for predicting opioid overdose

Year 2
-introduced new Deep learning algorithms perform well in low-cost platforms with high accuracy
-introduced novel deep neural networks to productively deploy Al based object recognition on mobile devices

-provided algorithmic analysis, competitive against large-scale deep networks, significantly reducing computational time and memory consumption
-evaluated on various applications in natural images and medical images
-lightweight model can achieve high performance on large-scale challenging natural and medical image benchmarking datasets
-developed new methods in predictive modeling framework for incorporating time-dependent features with a new method for deriving variable importance
-2 manuscripts under revision, expected to be resubmitted by end of April, 2022
-delivered oral and poster presentation at 2021 INFORMS Annual Meeting
-constructing optimization model to identify optimal size of time windows for prediction
-working on representation learning for our deep learning models

Year 3
-developed and published a # of novel and best-in-class self-supervised tools for a variety of application: 3D capsule networks for medical segmentation on less labeled data; multimodality multi-lead ECG arrhythmias classification; image deblurring; self-supervised domain adaptation deep learning method to deal with limited training data
-published a model for CLIP-assisted temporal self-attention for weakly supervised anomaly detection
-implemented a meta-learning of NAS for few-shot learning in medical imaging
-developed (2+1) D Distilled ShuffleNet: A Lightweight Unsupervised Distillation Network for Human Action Recognition

LP5: Harnessing Transaction Data through Feature Engineering
Year 1
-using MIMIC data
    -identified important features related to patient outcomes under ventilators
    -created descriptive statistics to be included in the feature set, and
    -presented preliminary result at 2020 INFORMS Annual Meeting

Year 2
-developed autoencoder method
    -invention in unsupervised and self-supervised deep learning methods contributed to tackle problem of large-scale dataset management and labeling in Big Data management
    -work provides an effective solution to solve Small Data challenges in Big Data era
    -introduced new combination approach between deterministic and probabilistic deep neural network
    -allows introducing a new powerful mechanism to reason knowledge representation in big data datasets
    -created new feature based on Tor network protocol
    -collects network traffic dataset over Tor network
    -classifies websites in real time
    -results in foundation technology for detecting websites that disseminate illegal contents
    -created low-dimensional feature using histogram entropy
    -used on malware datasets, Windows, Android, and loT malware
    -conducted study to acquire comparable results with small datasets to reduce cost of training machine learning models on huge datasets
    -performed a comparative analysis between convolutional neural network (CNN), residual neural network (RNN) and Vision Transformer (VT) using Ductal Carcinoma (breast cancer tissue images) dataset
    -assess suitability for adoption
    -VT model outperformed the CNN and RNN on different tasks achieving up to 93% accuracy
    -accepted paper on work
-introduced a pre-input layer to binary-decision-fusion neural network
-train layer for out-of-distribution cases adjusting feature values, without altering original training of network

Year 3
-developed a temporal clustering optimization model to determine the optimal patient-specific time-window size for irregularly-sample multi-variate transaction data

Conclusions and Recommendations
The Learning and Prediction research component is meeting most of its objectives. This component has transitioned MTPP and LTSM to a convolutional neural network (CNN) approach; curated dataset of sensor data, system attributes, and failure/repair data of over 8,000 oil/gas wells; investigated efficacy of group structure on generalized neural network (GNN) architecture with smallest finite simple nonabelian group A5 action of random and clustered synthetic small size data; introduced new Deep learning algorithms that perform well in low-cost platforms with high accuracy which significantly reduce computational time and memory consumption; developed autoencoder method invention in unsupervised and self-supervised deep learning methods contributed to tackle problem of large-scale dataset management and labeling in Big Data management.

Developed a novel self-supervised domain adaptation method in crowd counting and published two novel tools-Fairness Domain Adaptation (FREDOM) and Approach to Semantic Scene Understanding and self-supervised Spatiotemporal Transformers (SPARTAN) available on GitHub. The group has also published research papers in IEEE ACCESS detailing novel features for classifying malware using machine learning models as well as published an image preprocessing model at the Intl. Conference on Image processing Computer Vision & Pattern Recognition. Developed and published a # of novel and best-in-class self-supervised tools for a variety of application: 3D capsule networks for medical segmentation on less labeled data; multimodality multi-lead ECG arrhythmias classification; image deblurring; self-supervised domain adaptation deep learning method to deal with limited training data. Published a model for CLIP-assisted temporal self-attention for weakly supervised anomaly detection.

6. Education Component

Proposed: "*Developing a combination of model programs, degrees, pedagogy, and curriculum including a 9-week middle school coding block; a technical certificate, certificate of proficiency, and associate of science in data science; and a Bachelor of Science in data science with minors or concentrations. 2. Providing resources and training for educators including $5,000 Seed Grants for project-related Education & Broadening Participation; Career Development Workshops for project participants and educators; and K12 teacher professional development on data science topics. 3. Providing educational opportunities inside and outside the classroom for students. Undergraduate and graduate research assistantships in DART labs will be funded along with intensive data science and computing summer camps for undergraduates and research-based capstone projects and internships with industry partners. 4. Ensuring broad participation to impact the pipeline of data science skilled workers through Summer Undergrad Research Experiences in DART labs for underserved students, scholarships for underserved students to the Arkansas Summer Research Institute (ASRI); and by connecting students to opportunities through the Arkansas Center for Data Sciences (ACDS)."*

Findings:

Table of Selected Outputs of Component

Outputs	Year 1	Year 2	Year 3	Total
Publications	0	0	1	1
Joint Publications*	0	0	0	0
Presentations	0	5	2	7
Joint Presentations*	1	0	0	1
Workshop	0	4	1	5
Joint Workshops*	0	2	0	2
Proposals # ($)	0	0	0	0
Awards # ($)	0	0	0	0
Joint Proposals* # ($)	0	1 ($750k)	0	1 ($750k)
Joint Awards* # ($)	0	0	0	0

*Collaborative proposals/awards/publications with other DART components

Data Science Ecosystem: Model Programs, Degrees, Pedagogy, and Curriculum

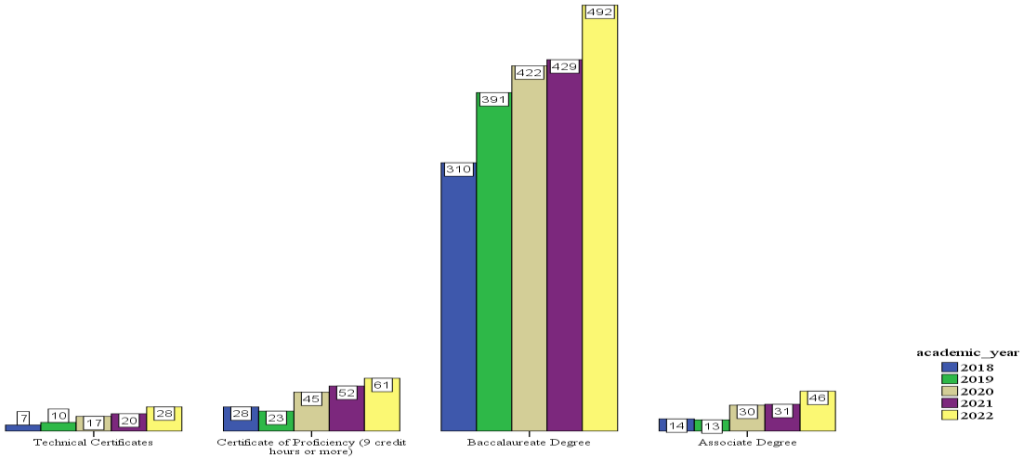
Year 1

-completed initial Middle School Coding Block Workshop with plan finalized and disseminated to stakeholders
  -5-year plan outlined for stakeholders during workshop November 2020
  -hosted approximately 50 attendees from 40 campuses and organizations in Arkansas
  -published curriculum information and plan on OneDrive site for participants at November workshop
-distributed institutional needs assessment to CS/DS faculty and IT support at all Arkansas campuses
  -approximately one-half of campuses responding
-UA Data Science program:
  -accepted its inaugural class with 45 students
  -approximately one-third are not calculus-ready,
  -one-third are "standard 8-semester plan", and
  -one third are transfer students from other majors or from other academic institutions
  -developed a suggested 6-semester plan for students who change their majors to Data Science and can be adapted for the second two years of 2+2 programs
  -representative on ABET's accreditation workgroup and is tracking requirements for readiness- program broadly distributed
-UCA developed a standalone BS in Data Science with concentrations in computer science, statistics, and business
  -program received academic approvals on campus and by UCA Board
  -submitted for review and approval by Arkansas Higher Education Coordinating Board
  -program specifics shared with cohorts at statewide meetings
-UAPB discussion on how to begin a Data Science program
-A-State developed BS in data science

-approved by Arkansas Higher Education Coordinating Board
-accepting students in Fall 2021
-Philander Smith College received approval for three courses:
-Intro to Data Science using python, Machine learning and Ethics in data science
-established partnership with IBM with 3 faculty completing Data Science training
-Shorter College finalized curriculum for two courses
-faculty are waiting to meet with Dean and Associate Dean for approval and implementation plan
-partnered with IBM with 4 faculty completing the data science and artificial intelligence badge programs
-completed 3 Data Science for Arkansas workshops
-attendance by post-secondary academic Arkansas institutions, Division of Higher Education (ADHE), Economic Development Commission (AEDC), and Center for Data Science (ACDS)
- shared via a OneDrive to participants, course materials, curriculum, adaptions and implemented curriculum
-identified "Opt-In" partners from the Research team

Year 3
-there are now three bachelor of data science programs in the state-UARK, A-State and UCA
-the first cohort of graduates matriculated from UARK in May 2023
-a key person has left the state and the team will be revising this strategy in their strategic plan

Figure 16
Data Science Graduates from Arkansas Institutions
by Academic Year and Degree Level



Data from the Arkansas Department of Higher Education shows the increase over the last five years of the number of degrees earned by students in Arkansas institutes of higher education. The largest increase (59%) was in the awarding of Baccalaureate Degrees (2018: 310: 2022: 492). Since a degree in Data Science was not well defined prior to DART these data include students who will become math teachers or other professionals, not necessarily data science professionals. As the project moves forward, we will refine these data to determine the true baseline for data science degrees in the state.

Conclusions and Recommendations
The Education research theme efforts in postsecondary education are moving forward. There are now three bachelor of data science programs in the state-UA at Fayetteville, A-State and UCA. The first cohort of graduates matriculated from UA at Fayetteville in May 2023. Five workshops on Data Science for Arkansas have been held during the first three years involving post-secondary institutions, ADHE and ACDS.

It is important to continue working with the cohort 1 institutions and begin work with cohort 2 institutions. The team should consider the recommendations by the EAB regarding courses to include in the Data Science degree programs in the state. An analysis of the historical data from ADHE on prior data science programs in the state may prove helpful in identifying those courses that have benefited graduates.

7. Workforce Development Component

Proposed: *"Provide K20 teacher and faculty opportunities for professional development spanning multiple disciplines*

Findings:

Outputs	Year 1	Year 2	Year 3	Total
Publications	0	0	0	0
Joint Publications*	0	0	0	0
Presentations	0	0	0	0
Joint Presentations*	1	0	0	1
Workshops	7	18	2	27
Proposals # ($)	1 ($1.8M)	0	0	1 ($1.8M)
Awards # ($)	0	0	0	0
Joint Proposals* # ($)	1 ($3.2M)	1 ($750k)	0	2 ($4M)
Joint Awards* # ($)	0	1 ($50k)	0	1 ($50k)

*Collaborative proposals/awards/publications with other DART components

Provide K20 teachers and faculty opportunities for professional development spanning multiple disciplines

Year 1

-launched March 2021 application for teachers EAST Initiative Annual Conference
-identified campuses that will submit faculty for Cohort 1 training:
  -Shorter College, PSC, UAPB, North Arkansas College, and Arkansas Tech University
-identified 5+ capstone partners from the research teams
-licensed new platform, UpSquad to serve as online community with teleconferencing and telework functionality

Year 2

-28 teachers have participated in a 3D printing workshop, 21 attended the Pi-Top workshop and 82 attended a leadership workshop
-offered 5 virtual Software Carpentry workshops attended by 195 participants (not all from DART)
-hosted a free workshop for 70 participants by NVIDIA Deep Learning Institute
-provided a Communicating Science to Legislators and Distilling Your Message workshops
-20 people attended a workshop on individual development plans
-hosted a DART pedagogy workshop

Year 3

- in collaboration with the EAST Initiative hosted an in-person workshop and two webinars followups for 28 teachers representing 22 school districts on 3D printers and another workshop on Pi-Tops for 17 teachers representing 14 school districts
-establishing an online program, LET'S Prepare for the Future in collaboration with EAST which will begin in August, 2023

Provide educational training opportunities inside and outside the classroom for students

Year 1

-provided training for graduate students and undergraduates throughout the year
-funded 10 underrepresented undergraduates for lab work during the summer

Year 2

-provided training and mentoring for 91 graduate students, 3 postdocs and 31 undergraduates
-hired 10 underrepresented undergraduate students for summer research

Year 3

-provided training for 117 graduate students, 4 postdocs and 50 undergraduates in DART
-PhD degrees were awarded to 7 DART research assistants and an additional 7 completed master's degrees
-Co-hosted with AR Department of Education and Division of Higher Education a Computing and Data Science Ecosystem workshop attended by 60 people across the state
-Co-hosted with AR Department of Education a virtual workshop for educators on K12 Data Science competencies and careers

Conclusions and Recommendations
The Workforce Development component is working closely with the EAST Initiative to embed Data Science into its programs. The Initiative is a 501-3C non-profit incorporated in 2006 and relies on a combination of legislative funds and corporate sponsors. Their mission is to provide K-12 students and teachers in Arkansas with access to technology and professional development to effectively integrate that technology into the curriculum. Partnering with EAST is a low cost and sustainable avenue for infusing data science into the educational fabric of public schools.

8. Communications & Dissemination Component
Proposed: "*Maintain interproject communication to accomplish milestones and relay updates."*

Findings
Year 1
-established Slack group for DART faculty who utilizes it for email for daily communication
-established DART GitLab with efforts being made to ensure cross-campus participation
-exploring an UpSquad license, a new telework and network community to be used for 2021 ASRI
-completed 11 monthly team meetings per 6 research teams
-hosted 5 DART Monthly Seminars to date with 2 additional webinars planned by end of Year 1
    -invited DART faculty, staff, and students to the webinar series
    -recorded DART webinars available on the DART website
-collected social media accounts and blogs of DART faculty for cross-posting
-established three listservs: one each for the DART SSC; DART Project Faculty and Staff; and DART students
-implemented ER Core site in August 2020 with participants onboard through March 2021
    -100% DART participants, paid and unpaid except advisory board members, provided accounts in ER Core
    - central office participating in ER Core Consortium and hired developers to make continuous improvements/upgrades to platform
    -70% of users attended at least one of the 5 webinar trainings from September to January
-published 9 peer-reviewed articles and juried conference papers included in the Scientific publication list and reported in NSF PAR
-2 statewide workshops for cohorts and waves.
-formed Science Journalism Committee

Year 2
-expanded communications with DART participants with a weekly digest email informing them of upcoming events, DART related funding opportunities and recent project accomplishments
-hosted weekly virtual office hours where a project leadership team member is available to answer participant questions
-held monthly seminars and monthly team meetings
-held a virtual retreat involving over 125 people, including 80 DART participants
-will hold first face-to-face All hands meeting before the end of project year

Year 3
-held its find in-person faculty and graduate student retreat at the Winthrop Rockefeller Center attend by over 125 people

Annual Retreat 2023 (excerpts from retreat evaluation report**)**
The retreat was held at the Winthrop Rockefeller Institute from January 12-14, 2023. Over 58 people attended during the two-day meeting; including faculty (36%); graduate students (45%); undergraduate students (12%); support staff (5%) and post-docs (2%). One-third (33%) of the attendees were from the University of Arkansas-Fayetteville; 12% from the University of Arkansas-Little Rock; 12% from the University of Arkansas Medical School; 10% from Southern Arkansas University; 5% from Arkansas State; 5% from AEDC and the remining from a variety of institutions.

A post meeting survey was distributed via an email invitation to participants the week after the retreat. The survey was taken by 26 of the 58 participants with a valid email address for a return rate of approximately 45%. Three-fourths or more of respondents rated the following aspects of the meeting and presentations 'Very Good' or 'Excellent': 'Presentation: "AR STEM Education & Workforce Development: An Update" by Tina Moore' (86%); 'Breakout Sessions: Strategic Planning with research teams' (83%)' ; 'Presentation: "AREON: Supporting Data Movement for Arkansas Research" by Elon Turner' (81%); 'Presentation: "Opportunities for Collaboration with the Semiconductor Industry & CHIPS+" by Walter Burgess, Matt Francis & Hugh Churchill' (78%) and 'Presentation: "Project Evaluation: Outputs and Metrics" by Kirk Minnick (77%).

The following are selected responses from the open response survey questions. The complete responses to these questions from survey respondents can be found in the Appendix.

What did you find MOST useful about the retreat?

- The 'networking' - the ability to have sidebar conversations with individuals that went beyond the normal presentations and group discussions. These types of conversations are harder to do in virtual events.
- Being able to discuss topics with other members where they are 100% available. The first two years have been difficult, imo, because people are not 100% available when participating in online meetings. Competing priorities and zoom exhaustion have made it difficult to be productive in meetings and have added a layer of uncertainty as to whether other members are engaged or disenchanted.
- The face-to-face meeting with other research tracks
- Meeting people in person, having conversations about their work and perspectives.

What did you find LEAST useful about the retreat?

- There was a lot that had to be crammed into one retreat, so it was hard to take deeper dives into each component, as well as spend quality time in discussion about planning revisions.
- Given the optional nature, many participants who should have been there were not there.
- A student poster session would have been good.
- It was too stretched out over 3 days. It could have all been accomplished in one full day. Too many people were missing.

Additional Comments:

- This was a very effective meeting that will significantly impact the course of the remaining DART activities.
- Overall thought that it was nicely prepared and well executed. Given the structure of this project, perhaps in the future the retreat should be focused on faculty as there was not much content or opportunity for the graduate students who attended.
- It was a great conference and very educational. The planning and location were also excellent.

Based on the responses on the feedback surveys, the retreat was a success at providing a venue for DART participants to learn about new opportunities and engage in strategic planning. Based upon the survey feedback, these participants would encourage project leadership to continue to provide this opportunity for:

- networking;
- strategic plan review;
- keep meeting content at a level that can be understood by someone unfamiliar with DART and
- facilitate face-to-face meetings between research tracks.

There were a few comments about the lack of representation from all the research groups at the retreat. In the future it should be mandatory that a minimum number of faculty from each component attend the retreat to have an effective planning session. Another model could be for the research components to conduct an initial planning session prior to the retreat; share their plan at the retreat and focus on integrating collaboration with the other components.

Annual Conference 2023 (excerpts from conference evaluation report)
This year the meeting was held at the Embassy Suites in Springdale Arkansas from April 20-21, 2023. Over 123 people registered for the two-day annual conference: including faculty/post-docs (33%), students (46%), Science Steering Committee (8%), Central Office (5%) and advisory board members (5%). Over one third (36%) of the registrants were from the University of Arkansas-Fayetteville, one-fifth (20%) from University of Arkansas-Little Rock, 10% from University of Arkansas Medical School and the remaining from institutions and organizations across the state.

A post meeting survey was distributed via an email invitation to the 123 conference attendees with valid email addresses the week after the conference. The survey was taken by 52 participants for a response rate of approximately 42%.

In general, respondents rated most aspects of the meeting as 'Very Good' or 'Excellent'. Three-fourths or more rated the meeting as ‘Excellent’ or ‘Very Good’ on the following: 'Promoting Inter-Institutional Collaboration’ (89%); ‘Promoting Within Project Collaboration’ (84%) and 'Promoting Inter-Disciplinary Collaboration’ (79%). Almost three-fourths (74%) of respondents used ‘Excellent’ or ‘Very Good’ as their rating for ‘Overall, it was well worth my time to attend the annual conference’.

The following are selected responses from survey respondents to the questions “What did you find MOST useful about the meeting” and “What did you find LEAST useful about the meeting?”. The complete set of comments can be found in the Appendix.

Most Useful:

- Networking, Collaboration, Student poster competition, Tech Talks delivered by Dr Xu, Trip to Crystal Bridges and the environment.
- Meeting with my theme team, a few cross-team discussions, and the student engagement and poster session.
- The poster session was the most useful since research from a variety of different disciplines was presented and gave the opportunity for presenters to practice science communication.
- What I found most beneficial was the ability to learn about emerging industry trends. The presentations on the use of data-informed decision-making inspired me, and I intend to pursue these ideas further in my work. Overall, the DART conference was an enlightening event for me.
- Found presentations on current topics like chatGpt , and industrial persons advices and learned more about DART .
- The poster session which involves students: It was great observing the young generations doing research and presenting their work. The panel discussion where panels share their experiences: their experience and sharing will inspire students to explore different choices for their future.
- The ability to talk in person and at the same time with all the key people involved in making decisions in the cyber-infrastructure of DART/ARP.
- Learning more about DART. As a first-year member, it was very helpful

Least Useful:

- For a "required attendance All Hands" meeting, there were *many* people who were not there who should have been there. That reduced the collaboration and discussions.
- Would have liked to see DART groups each present a summary on their state of affairs and progress
- Some of the faculties who presented about the state-of-the-art topics i.e. ChatGPT was not enough and I was under impression that they even don't know the bits and bytes of the topic.
- For social science researchers/participants, there was tendency to disconnect during more/hard scientific presentations, although some of the presenters did make good efforts to make their content understandable/relatable to all...
- Lack of technical exchange due to the weak agenda.
- Lack of communication & organization. Lack of overall participation (the event is mandatory but not everyone showed up). More time for dialogue conversations.

RECOMMENDATIONS

Based on the feedback surveys and observations by the evaluator, the annual conference was a success. The following are recommendation based upon comments made by survey respondents:

- Distribute the agenda earlier to the DART community.
- Include more sessions with students - maybe use the poster session to select a few students to give short talks about their research.
- Include a session led by young researchers (post-docs, PhD) from different thrusts.
- Include more technical and thrust presentations at the meeting organized by the SSC and Technical Director.
- Improve the poster competition process. Judges should complete their judging prior to the general audience and poster winners announced at the end of the poster session while the students are still there, and the attendees are viewing the posters.

Educate the public about DART accomplishments.

Year 1

-published basic project website presence for the project at https://dart.cast.uark.edu/

-details about each research theme and the faculty/graduate students,

-relevant documents, Strategic Plan, Arkansas S&T Plan and Annual Reports, and identified improvements

-published 2 AEDC blogs with 2 additional blogs end of year 1 at https://www.arkansasedc.com/news-events/arkansas-inc-blog

-hired communications intern at central office to assist in content generation and publicity of DART

-increased social media following on Facebook and Twitter by 7%

-formed Campus Communications Committee

-made initial contact with communications offices at most of the participating campuses

-host first meeting with Campus Communications Committee summer 2021

Year 2

-project website became live

-meet 5-year goals for social media followers on Twitter and Facebook

-met with campus communications committees

Year 3

-updated the project website to be more accessible by other researchers

Conclusions and Recommendations

The Communication and Dissemination component has transitioned well from virtual meetings to face-to-face. Both the annual retreats and annual conference are well attended and well received by participants. These events allow for networking among and between the research teams. The project website is used for communicating DART research to the broader research community and the public. The EAB recommended that the website be used to share research findings and datasets with other researchers.

9. Broadening Participation Component

Proposed: "*DART's leadership will take into consideration the institutional, gender, ethnic, and other forms of diversity of all participant groups and role types, including established panels. DART does not plan to hire new faculty or postdoctoral associates but will utilize the planned activities described here to broaden participation in the project. We will also encourage recruitment of women and underrepresented minority (URM) faculty and students for open positions. We will commit to the following targets for each role type:*

- *Faculty- 45% female, 10% URM*
- *Graduate Students- 50% female, 20% URM*
- *Undergraduate Students- 50% female, 40% URM*
- *Advisory Boards- 50% female, 20% URM"*

Findings:

Mentorship of students and early career faculty

Year 1

-filled 15 undergraduate (UG) student research assistantship positions
-filled 40 graduate (GA) student research assistantship positions

Year 2

-supported 40 graduate and 15 undergraduate students

Year 3

-project team has concerns about effectively implementing these goals in the current political environment in the state
-identifying potential partners and organizations that are not constrained or influenced by new state regulations on diversity

Research Seed Grants Program

Year 1

-first round of seed grants offered in October 2020 with two awards of $5,000 issued to:
-Arkansas Regional Innovation Hub for a virtual field trip project entitled "STEM Saturdays", and
-Henderson State University STEM Center for a project entitled, "ConneCTED: Developing Communities of Practice with an Emphasis on Computational Thinking and Engineering Design"

Year 2

-second round of seed grants awarded over $752k to 10 projects
-ATU: "Development of Interdisciplinary Research Collaborative to Provide Datasets in Support of Education Research in Data Science" (ED)
-UARK: "Toward Fair and reliable consumer acceptability prediction rom from appearance" (LP)
-UAMS: "Geospatial Data Science in Public health: Interinstitutional educational collaboration to enhance data science curriculum in Arkansas" (ED)
-UARK: "Interpretable Multimodal Fusion Networks for Fault Detection and Diagnostics of Two-Phase Cooling Under Transient Heat Loads" (LP)
-UAMS: "Piloting Bid Data Science in Arkansas Middle School Classrooms" (ED)
-UARK: "Machine Learning-based emulation and prediction in ensembles in disordered photocatalytic composites" (LP)
-A-State: "AgAdapt: An evolutionarily-informed algorithm for genomic prediction of crop performance in novel environments" (LP)
-UALR: "Machine Learning for Predicting Refugee Counts" (LP)
-UCA: "Crying out Data Science in the center of Arkansas: Invitation to High School students to the World of Data Science" (ED)
-PSC: "Generating Big Radiogenomic Data of Cancer Using Deepfake Learning Approach"

Year 3

-third round of seed grants awarded over $505k in 2022 to 6 projects
-A-State: "MoDaCoM-TL: Model and Data Compatibility Metric for Transfer Learning" (LP)
-UAMS&A-State: "Smart curation and deep learning-based enhancement of social risk data" (DC/LP)
-UARK: "AI-Supported Cyberinfrastructure for Scalable Flood Resilience Assessment" (CI)

-UAMS: "Machine Learning Approaches for Remote Pathological Speech Assessment for Parkinson's Disease" (LP)
- Progress on first round seed grants listed below
  - 'AgAdapt: An evolutionarily-informed algorithm for genomic prediction of crop performance in novel environments'
    -trained 4 UG students. Three of whom are URM in STEM and majoring in Math or Computer Science
    -selected for National NSF I Corps award of $50k to investigate commercialization of our technology
  -'Machine Learning-based emulation and prediction in ensembles in disordered photocatalytic composites'
    -developing neural network emulators for predicting light concentration
    -developed an approach based on PCA to predict where the light will concentrate on omission glass
  - 'Development of Interdisciplinary Research Collaborative to Provide Datasets in Support of Education Research in Data Science'
    -offered 8 sessions of Data Science workshops in 2021 and 2022 through Webex to student, faculty and staff at ATU
    -trained 10 UG students on R code and presentation skills
  - 'Generating Big Radiogenomic Data of Cancer Using Deepfake Learning Approach'
    -support 5 UG students, all URM
    -taught deep learning approached to students
    -developing a Generative Adversarial Network GAN to generate deepfakes
- 'Piloting Bid Data Science in Arkansas Middle School Classrooms'
    -developed data science curriculum for use on chrome books
    -implemented curriculum in 4 school districts involving 4 schools, 5 teachers, 32 classrooms and 770 students
    -collecting pre-post student data to assess impact
- 'Interpretable Multimodal Fusion Networks for Fault Detection and Diagnostics of Two-Phase Cooling Under Transient Heat Loads'
    -developed a machine learning framework to determine the heat flux during pool boiling using acoustic signals captured through a hydrophone
    -the team applied and participated in the National NSF I-Corps program
-'Crying out Data Science in the center of Arkansas: Invitation to High School students to the World of Data Science'
    -provide 15 high school students with a one-week workshop on data science
    -presented the workshop the Arkansas Council of teachers in Mathematics
    -trained 1 graduate and 1 UG student who helped with the workshop

Summer Undergraduate Research Experienced (SURE) Program

Year 1

-supported 12 underrepresented undergraduate and 1 high school students in DART labs

Year 2

-intend to fund an additional 15 undergraduate students in the summer of 2022

Year 3

-none reported

Broadening Participation Seed Mini-Grants

Year 1

-none reported

Year 2

-Awarded 7 mini-grants a total of $15.5k
    -UCA: "Integrating Data Science to Rethink Mathematics and Science"
    -Harding: "Robots, Rocketry and Programming summer Camp"
    -Ozarks Unlimited: "K-5 STEM Integration: Designing Authentic & Meaningful STEM"
    -Northwest Arkansas ESC: "K-5 STEM Integration: Designing Authentic & Meaningful STEM"
    -Northcentral Arkansas ESC: "K-5 STEM Integration: Designing Authentic & Meaningful STEM"
    -Dawson ESC: "K-5 STEM Integration: Designing Authentic & Meaningful STEM"
    -Southside School District: "Southside Summer STEM Institute"

Year 3

-none reported

Arkansas Summer Research Institute (ASRI)
Year 1
-none reported

Year 2
-40 of the 100 student applicants recruited for attendance at Arkansas Summer Research Institute, ASRI
-42 undergraduates completed the ASRI virtually, which also involved 53 presenters and panelists from DART institutions

Year 3
- the ASRI was attended by 102 undergraduates. Over half (52%) of the participants reported as Male, 45% Female and 3% Non-binary or non-conforming. The ethnic distribution was over one-third (36%) Asian; more than one-fourth (28%) Black; more than one-fifth (23%) Hispanic; 17% Black; 9% Multi-ethnic and 6% Other. Almost three-fourths (73%) of student participants reported attending an institution in Arkansas.

ASRI: Excerpt from Year 2 evaluation report
The ASRI was held virtually using Zoom for two weeks from June 6-17, 2022. The ASRI used a variety of virtual tools, including Zoom, Google Suite, Peardeck, Scoreboard, Flipgrid, Magoosh and Labster. The ASRI was attended by 102 undergraduates. Over half (52%) of the participants reported as Male, 45% Female and 3% Non-binary or non-conforming. The ethnic distribution was over one-third (36%) Asian; more than one-fourth (28%) Black; more than one-fifth (23%) Hispanic; 17% Black; 9% Multi-ethnic and 6% Other.

Almost three-fourths (73%) of student participants reported attending an institution in Arkansas. More than one-third (43%) of the students are attending an institution in the University of Arkansas system; UA: Fayetteville (31%); UA: Little Rock (11); UA: Pine Bluff (1%) and University of Arkansas for Medical Science (1%). Other institutions in the state with attendees are Arkansas Tech University (7%); Arkansas State University (5%); University of central Arkansas (5%); Henderson State University (2%); Northwest Arkansas Community College (2%); Philander Smith College (2%); Bentonville High School (1%); Harding University (1); Hendrix college (1%); Ignite Professional Studies (1%) and Southern Arkansas University (1%). Students attending colleges outside Arkansas were also able to participate, since the institute was held virtually. These included Mississippi State University (19%); University of Oklahoma (2%); Southern University and A&M College (2%); Wake Forrest University (1%); Pennsylvania State University (1%) and University of Memphis (1).

The percentage of students rating as 'Excellent' their overall daily ASRI experience ranged from less than half for Week 1: Monday (45%) and Tuesday (43%) to three-fourths (75%) for 'Week 2: Friday. Overall daily ratings with a combined 'Excellent' and 'Very Good' all exceeded 80% for the two weeks.

In addition to the student participants, the ASRI involved 52 presenters and panelists, which included graduate students, faculty, staff and entrepreneurs. About one-fifth (21%) were from the University of Arkansas at Fayetteville; 14% from Arkansas Tech; 14% from the University of Arkansas for Medical Science; 8% from the Arkansas School for Mathematics, Sciences, and the Arts; 8% from Arkansas State University and 10% representing industry.

Presenters were involved in the ASRI in multiple roles. Half (50%) were session presenters; 42% were 'helping students prepare presentations'; 19% were 'Panelists'; 15% were 'Facilitating Networking and breakout rooms' and 15% were 'Organizers'. Four-fifths (81%) of the presenters responding to the survey were 'Faculty'; while 2 (8%) were 'Staff'; 2 (8%) from 'Industry' and 1 (4%) was a 'Volunteer'.

RECOMMENDATIONS
The ASRI students were asked: "Do you have any ideas how to make the overall experience even better?"
- Create an Excel track. In my discipline of economics, we use Excel (almost solely at MSU) for data analysis. I'd love to learn more about Excel and explore the full capabilities of its' power to use in my time as a student and as a professional.
- I think, for the future, there should be more sessions that help prepare students for the job market as there are alot of graduates, grad students, and post-grads that seem to have attended this year.
- I think ASRI should be extended another week. I appreciate the short time period because non-traditional students are able to complete the experience. However, the presentations do feel a bit rushed. I felt as if I was still learning techniques to put into my presentation the day before its due date. Maybe the first two weeks can be devoted to presentation techniques, data analysis, etc. Week three could be a hybrid of more information but mostly breaking down presentation information.

The ASRI presenters were asked: "What would have made the ASRI experience better for you?"
- From an organizer's point of view, it would be better to expand it to three weeks and not have sessions all day of every day to allow more work on their projects. We would have been able to work with them more, as would their mentors.
- I would have liked the opportunity to engage more with the ASRI students. Getting to know the students helps a bit in determining what is most helpful to share during a subject matter panel.

Conclusions and Recommendations
The Broadening Participation component is achieving many of its goals by providing professional development for undergraduate and graduate students who tend to be female and URM. The group has also awarded 2 broadening participation mini-grants of $5k each in Year 1 and 7 mini-grants a total of $15.5k in Year 2; hosted the ASRI for two years; supported undergraduates and graduate students each academic year and supported URM students each year in the SURE program.

The evaluation of the ASRI found that this program was engaging, educational, inspirational, comprehensive, rewarding and fun for the participants. Faced with the challenge of Covid-19 the leaders did a lot of planning and hard work to pivot from a face-to-face Institute to one that was totally virtual. They did not try to replicate the face-to-face training but instead thoughtfully designed a wholly new program that maximized the strength of a variety of virtual tools available to them and the students. While the ASRI could have been held face-to-face in Year 3, it was decided to continue the virtual program as it allows more student participants as well as more engagement by faculty.

This group needs the support of the research themes to achieve its diversity goals for: Faculty (45% female, 10% URM); Graduate Students (50% female, 20% URM) and Undergraduate Students (50% female, 40% URM). According to the self-reported participant data for Year 3 the project shows: Faculty (31% female; 4% URM); Graduate students (34% female; 12% URM) and Undergraduate students (45% female; 46% URM).

10. Management, Evaluation, and Assessment Component

Proposed: *"A senior management team oversee research activities within and across the topic areas. It will be the responsibility of this team to ensure that collaborative activities are ongoing, productive, and fall within the defined research goals of the project…. DART will have an Industry Advisory Board (IAB), an External Advisory Board (EAB), and an external evaluator/consultant. The IAB members will serve as an intermediary between academia and industry. The IAB will include representatives from Arkansas industry sectors who will be impacted by DART research….The program leadership, External Evaluator, and EAB will monitor and assess program activities. Project leadership is responsible for ensuring that data is collected on milestones, participants, proposals/awards, publications, and other project outputs, as well as implementation of recommendations from the EAB and evaluator. The EAB will assess overall results and identify areas of concern and new opportunities for DART as they arise. They will conduct an annual site visit and produce a report that will be forwarded to the NSF Program Officer…."*

Reverse Site Visit

The Reverse Site Visit (RSV) panel met virtually because of COVID with the DART project team over zoom video/audio session on April 11-12, 2022. The DART management team presented their accomplishments and progress virtually to the panel.

While the RSV identified many areas of strength for DART, they also noted some areas for improvement. The panel's overall recommendations are listed below.

Recommendation 1. Website – Provide a plan for developing website sections listing the DART project's research publications and offering the project's datasets to the research community at large.

Recommendation 2. Cyberinfrastructure – (i) The panel did not find satisfactory details describing the architecture of the Arkansas Research Platform. Provide a detailed diagram showing the participating organizations, the connections, and the physical cyberinfrastructure. (ii) A significant sociotechnical challenge is federated identity management across participating institutions. Provide a workgroup document outlining the detailed needs assessment and the solutions for the sub-challenges identified for federated identity management. (iii) Provide a process document for the use of cyberinfrastructure during allocations, onboarding, and supporting user data operation needs. This can be coupled to an analysis of the state-scale common application themes based on interests, strengths, and needs. (iv) Implement a cyberinfrastructure risk register, with mitigation measures, and a monitoring dashboard to track and visualize cyberinfrastructure usage by various user groups.

Recommendation 3. Project management – Explain how the technical aspects of the DART project are being managed using the methods of formal project management, and who is qualified and responsible to conduct the above level of management.

Recommendation 4. Assessment – Explain how the evaluation plan provided by the project's external evaluator analyzes how research efforts, educational outcomes, seed grants, and student awards contribute to the project goals of enhanced research capacity.

The EAB has met two times during the project, once briefly during the first year virtually and a second time during the Year 2 DART Annual Meeting in Little Rock. Excerpts from the second report are presented below.

1. Coordinated Cyberinfrastructure

The Coordinated Cyberinfrastructure Team has a set of challenging aspects to address. First and foremost is the federated identity and access management of the respective campus research computing resources within the Arkansas Research Platform contributors.

Parallel to the federated access POCs, the EAB suggests the team explore methods to standardize the environment for collecting metrics from the ARP clusters. As ARP is further formalized and leveraging the respective computing clusters, collecting and reporting metrics across the participants becomes a priority to demonstrate the value. Open XDMoD would be a natural fit.

The EAB suggests the campus CISOs or designees of the participating campuses have a series of focused conversations and documentation of the practices and policies in place. In addition, they should include discussions with other regional campuses to compare existing multi-institutional research computing sharing practices. EAB suggests leveraging the SHARP-CI team.

2. Data Life Cycle and Curation

The EAB suggests that the team share their findings with the DART team and the broad scientific community. The idea of holding a team retreat this fall to revise the DC strategic plan is very good. It has the potential to bring the whole team on the same page in terms of language and terminology, to make a cohesive goal for this component, and the adoption of the proposed technologies.

Since facilitating the collaboration among multiple campuses in Arkansas is one of the goals of this project, the EAB suggests exchanging Ph.D. students among different campuses during the summer might be a good idea. Also, during the retreat or annual conference, bringing diverse groups together to discuss the proposal submission plan for the next year might be a good direction too. The EAB is delighted to see that the industry collaborations in this team are strong and continue to grow.

3. Social Awareness

The social awareness (SA) team has done a great job writing literature review papers and survey papers on privacy, statistical fairness, and causal fairness. They have developed various algorithms to provide privacy preservation, fairness, safety, and robustness of data analytics, data collection, data sharing, and decision making, with multiple publications in leading conferences and multiple awarded grants and training of numerous graduate students. All seven projects have been conducted as scheduled. It seems that SA1 and SA2 have been completed. The EAB would like to see the SA team plans to expand the activities, have better communication of SA work and results, and invite a social scientist member to the team.

4. Social Media & Networks

The EAB believes that good progress has been made and that this component is meeting most of its objectives. The SM team has determined key features and software design document for the cyber social network platform; developed, tested, and deployed a data collection framework with real-time dashboard to monitor progress with alerting capabilities; revised the taxonomy to characterize OIE based on social media platforms; studied cyber campaigns and characteristics of platforms and involved information actors and selected Hurricane Harvey to represent a large-scale and geographically widespread disaster scenario. Research into the auto-annotation of multimedia data goal appears to be trailing the progress being made in the other goals/objectives, as noted by the external reviewers. Please address the concerns raised by the evaluator, including “obtain and index content types for at least two disaster scenarios” and “developing GIS system to display real-time road status inputs” have not been done yet.

The EAB suggests the project team consider the following questions raised in the RSV report as potential areas of interest for the next step, including: (i) how to incentivize users to provide trusted information; (ii) how to develop trust models and handle biases; and (iii) how to reliably assess information-driven context or situation awareness leading to safety and security. The number of papers generated in this component is beyond expectation; however, the EAB suggests that the team keep improving the quality of the work by publishing in top conferences in the field. Targeting the conferences listed on ankings.org might be a good start. Also, as one of the co-leaders of the team left the university, the EAB suggests that the project leadership take action as soon as possible to revise the SM strategic plan in consideration of the new team member's expertise.

5. Learning and Prediction

The EAB noted that the team implemented the EAB suggestions from last year with regard to greater elaboration on its achievements. The EAB suggests connecting Learning and Prediction more closely with other project areas, such as Social Awareness. It is indeed exciting to note that this component has grown to cover a wide range of important activities that are core to the DART mission. Therefore, the EAB suggests that the team consider structured/facilitated ideation workshops involving the different LP project members with project members from other components (e.g., Data Curation, Social Awareness, Social Media

and Networks, Education) to facilitate cross-fertilization of ideas and possibly address the RSV concern of impact and outreach both within DART and its industry partners and across the broader community.

6. Education

The EAB noted that the DART educational team reports significant postsecondary initiatives across six campuses spanning both community colleges and universities within the state. Additionally, some key efforts in the K-12 space were also noted. Finally, the DART ED team reports several strong government, industry, startup (including through the NSF I-Corps program) and non-profit collaborations which can help improve educational outcomes and opportunities for graduates.

The EAB reiterates its suggestion from last year to emphasize the importance of continuous evaluation of programmatic elements – this becomes increasingly important as the program's first graduates join the workforce next year. The EAB was pleased to see examples of ethics in data science curriculum development and would like to see more details on this element in next year's report. The EAB does have several suggestions for the DART team to consider. First –given the impressive array of collaborations identified – board members suggest leveraging these industry and government collaborations to help amplify and highlight the importance of data science careers in local industries. Second, such connections can be leveraged in capstone projects within the curricular infrastructure being proposed. Third, research projects from other components of DART (e.g., seed grant recipients) can serve be leveraged in such courses and can serve as exemplars of research in pedagogy. Finally, as the first graduates come through the program, the DART ED team needs to start to think about holistic programmatic outcome assessments.

7. **Workforce Development and Broadening Participation**

One possible way to attract more attention to the SURE program could be to reach out to the smaller universities in Arkansas, for instance, the PUIs, and the HBCUs. These are more likely to have students with less access to research opportunities in areas covered by the DART project. Thus, working with faculty at these smaller institutions/universities and then reaching out to their students could improve awareness of the SURE opportunity for such students. Also, providing the SURE program as a summer internship for the target students might help enhance interest in the program. To encourage faculty in the smaller universities to be more involved with the SURE program, DART could also provide incentives for faculty members that take up SURE students, for instance, by providing summer support to the faculty. Since most of these faculty members are involved in teaching to cover their summer, this could also help them put more effort into their own research activities, especially during the summer.

8. **Communication and Dissemination**

Communication and dissemination are extremely important, and maintaining internal communication across campuses and research is challenging. The EAB provides a variety of possible technology stacks that have been used in different campuses successfully. The two main ecosystems are Microsoft Teams and Open Source.

MS Teams can serve as a highly integrated platform for chat, conference calls and file transfer. The main challenges people have faced are permissions and user management. If all campuses in DART already have Microsoft Teams subscriptions, this may be a more or less complicated option. One problem for MS Teams may be centralized user management based on AD. This may simplify the setup here if this is set up at the central office level. MS Teams is not as powerful for real-time communications, which may be a desirable feature for close collaboration between DART members and groups.

An alternative tech stack is based on separate, freely available components. The main piece that we have seen being used is a discord for real-time communications, which is already used by many contributors already and makes it fairly simple to add a DART-specific group. For longer-term communications, we have seen good success with using web fora like phpBB or similar systems for recorded conversations. For maintaining shared definitions or onboarding documents, a shared document platform like a wiki like MediaWiki (or many others) has proven useful. Other shared document options are possible, but Wikis seem common enough for people to accept them fairly easily and successfully. This will need some seeding, possibly from the central coordination group for DART, but once the basics are provided, Wikis tend to grow well.

For maintaining more project progress-related communications, workflow tools like Trello or Asana can easily maintain tasks and progress histories in an easy-to-understand and digest format. When used consistently, they can significantly simplify new user onboarding. Neither is Open Source, but they have free usage tiers for smaller projects. There are Open Source alternatives like taiga.io and others if needed.

The EAB believes that different technological solutions will enable effective post-covid communications and would encourage the project team to pick the easiest to integrate and closest to what is being used in the different groups already. We have had limited success with adding too many new components to existing environments unless no tech stack is overlapped between the admittedly pretty heterogeneous and fairly large number of groups in DART. We would encourage the central group to survey tools already in use and find the intersection between all the DART groups. For dissemination, the project website is the primary tool, and it already does so very well. Including a section that clearly highlights project outcomes would be beneficial.

Conclusions and Recommendations: The Management, Evaluation, and Assessment Component through its project leadership has worked to ensure that the research teams have benefited from the feedback provided by both the EAB and RSV panel. The leadership provided time at the annual retreat for the external evaluator to present the findings from these two external reviews and to encourage the research groups to revise their strategic plans based upon these findings. The IAB has been less involved in providing feedback but has been involved in sharing industry needs with the DART project.

DART has more than tripled the number of research groups that need to be monitored with the awarding of 16 seed grants. In addition, the project has also awarded broadening participation seed grants to several institutions. The leadership team may want to consider expanding the central office staff to include another professional to help oversee these additional project activities and ensure that these awards fulfill their promises and contribute to the overall mission of DART.

Strategic Plan

A virtual strategic planning meeting was held in August 2020. The final strategic plan was submitted to NSF in December 2020. The plan provides metrics with baselines and 5-year targets to determine whether the project is on track. The strategic plan metrics for the entire project can be found in the Appendix of this report. A summary of the strategic plan metrics is presented in Figure 17.

Figure 17
Strategic Plan Metric Summary Table

	Metric Status				
Strategic Priority Area	**#**	**Met/ Exceeded Target**	**On Track**	**Delayed**	**Not Met**
Cyberinfrastructure	11	8 (73%)	1 (9%)	2 (18%)	0
Research	9	6 (67%)	2 (22%)	1 (11%)	0
Education & Workforce Development	21	2 (10%)	15 (71%)	4 (19%)	0
Communication	10	3 (30%)	6 (60%)	1 (10%)	0
Broadening Participation	15	1 (6%)	10 (67%)	4 (27%)	0
Overall	66	20 (30%)	34 (52%)	12 (18%)	0

Overall, DART has ‘Met/Exceeded Targets’ in 30% of its strategic plan metrics, are ‘On Track’ to meet in 52% and are ‘Delayed’ in 18% of the strategic plan metrics.

The Cyberinfrastructure Component has ’Met/Exceeded Targets’ in 73% of its metrics, are ‘On Track’ to meet in 9% and are ‘Delayed’ in 18%. Those metrics which are delayed are: ‘Instructors trained in software carpentry’ and the number of ‘Online workshops’.

The Research Component has ‘Met/Exceeded Targets’ in 67% of their metrics, are ‘On Track’ to meet in 22% and are ‘Delayed’ in 11%. The metric which is delayed is the proposed number of ‘Applications and Platforms’ in the various research thrusts.

The Education and Workforce Development Component has ‘Met/Exceeded Targets’ in 10% of their metrics, are ‘On Track’ to meet in 71% and are ‘Delayed’ in 19%. The metrics which are delayed are: number of ‘Seed grants awarded to K-12 teachers’, ‘number of industry internships awards’, ‘% internship students supported who are URM/female’ and the development of ‘Industry internship evaluation form’.

The Communication Component ‘Met/Exceeded Targets’ in 30% of their metrics, are ‘On Track’ to meet in 60% and ‘Delayed’ in 10%. The metric which is delayed is the ‘# of DART topical webinars held’.

The Broadening Participation component has 'Met/Exceeded Targets' in 6% of its metric, are 'On Track' in 67% and 'Delayed' in 27%. Those metrics which are delayed are: 'Provide training to mentors on how to use IDP and to be a mentor', 'Early career faculty complete IDP templates pre and post', 'Award maximum of $8k per SURE faculty' and 'SURE Students complete IDP template at beginning and end of experience'.

Summary and Recommendations

DART has involved 261 direct participants in the project so far. Almost half (44%) of the project participants are 'Graduate students' while another one-fifth (20%) are 'Undergraduates'. Less than one-third (31%) are 'Faculty participants' with Post Docs representing 1% of the total number of participants. The remaining participants were 'Support staff' (4%).

DART involved 261 participants in program Year 3, an increase of 45 (21%) from Year 2. In Year 3, undergraduate students continue to be the most ethnically diverse group of participants with 46% self-identifying as an ethnic minority, while 12% of graduate students and 4% of faculty self-identify as an ethnic minority. No post-docs or support staff self-identified as an ethnic minority. The overall ethnic diversity across all participant groups was 16%.

Female diversity is much higher with more than one-third (36%) of all participants self-reported being female. The participant groups with the most females were post-docs (75%); support staff (60%); undergraduates (45%); graduate students (34%) and faculty (29%).

Over the three program years DART has involved 282 unique individuals: almost one-third (31%) faculty; over one-third (42%) graduate students and more than one-fifth (22%) undergraduates. Across all participant groups 17% self-identified as an ethnic minority and one-third (37%) self-identified as female.

The institutions with the largest number of participants involved all project years are: University of Arkansas-Fayetteville (31%), University of Arkansas at Little Rock (28%); University of Arkansas for Medical Sciences (7%); Southern Arkansas University (7%); Arkansas Tech University (5%) and University of Central Arkansas (5%). The project has also involved participants from other higher education institutions in the state, including many smaller institutions, such as Arkansas State University (4%); Philander Smith College (5%); University of Arkansas at Pine Bluff (3%); Shorter College (2%) and North Arkansas College (<1%).

Project participants reported 28 external collaborators during the first three project years. More than one-third (43%) of the collaborators are at 'academic research institutions', while more than one-tenth (14%) are from 'primarily undergraduate institutions' and 'industry' (14%). Almost two-thirds (64%) of the external collaborators reported are in Arkansas; while one-fourth (29%) are collaborators located outside Arkansas but within in the US. Only 7% of the external collaborators are from outside the US.

While it is important to increase the number of collaborations with other researchers, it is equally important to involve collaborators from a variety of institutions. DART external collaborators are at 25 different institutions; one-third (36%) are 'Research Institutions', while one-fifth (20%) work at 'Industry/Business' and less than one-fifth (16%) are at 'Primarily Undergraduate Institutions'. Over half (56%) of the external collaborating institutions are in Arkansas.

Over $1.2 million has been expended or ordered during the project to purchase computer infrastructure at three institutions. In Year 1, project funds were used to purchase $650k of infrastructure at the University of Arkansas at Fayetteville. The equipment purchased included DELL Fiber Splitter cables, PowerEdge XE8545, Power Edge R7525, Server, NVIDIA Ampere A100 649,607.18 and a 40-port Mellanox Quantum QM8790. This equipment is part of the DART CI Plan that increases additional hardware needed to move

pinnacle out from behind the firewall. Also, in Year 1 data storage servers allowing data sharing among DART researchers were purchased for $24k at the University of Arkansas at Little Rock.

In Year 2, $496k was used to upgrade the research backbone at University of Arkansas Medical Sciences to collaborate with ARE-ON and extend service to the University of Arkansas at Fayetteville.

In Year 3, no project funds were budgeted or used to purchase equipment or infrastructure.

Ninety-six proposals requesting over $87 million have been submitted by DART participants. As of February 29, 2023, 54 proposals have been funded for a total of $31.7 million, while seventeen are still pending a funding decision. The funding agencies where most of the proposals have been submitted are NSF (34%) and NIH (13%), while the most award dollars have come from NSF ($8.7M or 27%); US Department of Defense ($7.7M or 24%) and US Office of Naval Research ($4.5M or 14%).
NSF funding in the state has increased from $14 million in 2001 to $32 million in 2022, in 2022 adjusted dollars. While the years 2010 to 2013 showed marked increases in award dollars to the state, this was the result of the American Recovery and Reinvestment Act. A similar increase in Federal funding occurred in 2020 and 2021 with the Covid-19 Stimulus funding. However, Arkansas is on track to maintain a high level of award dollars coming into the state.

Since 2001, Arkansas has received over $493 million dollars in NSF funding. Most (43%) of this funding has been in the form of research awards totally $215 million; co-funded awards (both for research and EHR) have amounted to $102 million (21%); EPSCoR awards (Tracks1, 2 and 3) have amounted to $94 million (19%) and $82 million (17%) has been awarded for EHR proposals.

NSF funding for Research, excluding EPSCoR co-funding and Track1, has increased from $5 million in FY 2001 to over $15 million in FY 2022, an increase of 200%. In addition, Education and Human Resource funding from NSF has doubled from $2 million in FY 2001 to $4 million in FY 2020.

DART participants reported making 143 presentations, posters, and invited talks during the first three program years. Almost half (43%) were ‘Presentation/Talks’, while one-third (33%) were ‘Invited Speakers’ and one-fifth (21%) were ‘Posters’.
Researchers reported 143 peer reviewed publications in the first three years of the project. Over one-fourth (26%) reported receiving primary support from DART. Some of the journals in which DART researchers have published included: BMC Bioinformatics; Infection and Immunity; Metabolites; Microbial Genomics and International Journal of Advanced Computer Science and Applications to name a few.

An analysis of these peer reviewed publications from the first three years was conducted using Clarivate’s Web of Science. A total of 101 or 71% of the 143 publications were able to be matched with those in the citation index. As expected, the majority (62%) of DART publications were categorized in the Computer Science content area. Other content areas these publications covered include Mathematics (23%); Engineering (22%); Mathematical Computational Biology (20%); Communication (18%) and Genetics Heredity (18%).

These 101 publications were cited in 412 articles for an average of 4.3 times per publication. The H-Index of the DART faculty was 9. The H-index or Hirsch index is a metric for evaluating the cumulative impact of an author or group of authors scholarly output and performance. It measures quantity with quality by comparing publications to citations. A group of researchers with an h-index of 9 have published nine papers that have been cited at least nine times by other scholars.

There have been six reported patent disclosures filed since the start of the project, which have all been in the Learning & Prediction component:

- Multiple Camera Multiple 3D Object tracking on the Move for Autonomous Vehicles (Khoa Luu)
- Self-supervised Domain Adaptation in Poultry Counting in the Wild (Khoa Lee)
- Smart Insect Control Device via Artificial Intelligence in Realtime Environment (Khoa, Tanh-Dat Truong)
- Sensor-based Smart Insect Monitoring System in the Wild (Khoa Luu, Tanh-Dat Truong)
- Artificial Intelligence and Vision-Based Broiler Body Weight Measurement System and Process (Thi Hoang Ngan Le, Khoa Vo)
- Chicken Processing Plant with Automated Computer Vision (Thi Hoang Ngan Le, Khoa Vo)

Overall, an estimated 1,862 people have been involved in one or more external engagement activities supported by DART during the project. More than two-thirds (70%) were K-12 students reached directly. Two-thirds (44%) of the students reached directly were female and 30% were an underrepresented minority in STEM.

The Coordinated Data Science Infrastructure component has been successful at acquiring and installing hardware, but less so in providing access to that hardware for users.

The team has purchased and installed: 20 nodes dual AMD 7543, 1024GB, NVMe local drive, single PCI 40GB A100GP; 4 nodes dual AMD 7543, 1024 GB, NVMe local drive, four SXM 40 GB A100 GPU and 100gb switch for UAMS. The visualization for complex data in diverse data-analytics application domains has been delayed by loss of personnel.

The issue of federated identity management across participating institutions is being addressed by using UA Systems Bridge to automatically authenticate users. The team still needs to fully automate this process without compromising security. Workshops on using an interactive shell to access Pinnacle are helping to increase the use of these computer resources by faculty and graduate students. A one-on-one connection between IT professionals and researchers seems the best approach for increasing HPC usage.

The evaluator has some concern that the goal on visualization for complex data will be able to be completed. The loss of personnel and the additional work needed to implement federated identity management may impact its completion.

The Data Life Cycle and Curation component is meeting most of its objectives.

Faculty have a prototype version of the washing machine; automated data cleansing to include data corrections based on record-to-record comparisons with blocks and clusters; developed computational framework integrating multi-layer genomics data to identify transcriptome and pathway dysregulations in autism spectrum disorder; investigated the expression alterations of survival-related genes in various immune cell types when combining breast cancer bulk and single-cell RNA sequencing data; established a computational workflow of several combined machine learning approaches to identify biomarkers for both prostate cancer using metabolomics data and chemotherapy-induced cardiotoxicity among breast; downloaded more than 300k bacterial and archaeal genomes from the NCBI and complete set of genomes from Integrated Microbial Genomes and Microbiomes project; developed and published a program, ProdMX, to speed up genome comparison more than a million-fold compared to traditional alignment methods; and published a paper giving an overview of multi-omics approaches in the journal, “Molecular Omics” among other accomplishments.

It is commendable that the team is incorporating suggestions made by the EAB and RSV into its work. It is not clear that other components in DART or others in the data science community are using any of the algorithms or programs/processes developed by this research group. Recommend that this group do more to advance the use of their algorithms, perhaps holding a workshop for DART faculty and graduate students and tracking the use of DC algorithms in the broader data science community. The DWM programs may be of interest to those college students in some of the new data science programs being developed around the state.

The Social Awareness component activities and progress have been impacted by the loss of key personnel. Prior to Year 3 this component had researched representative algorithms, using threat models from four aspects: adversarial falsification, adversary’s knowledge, adversarial specificity, and attack frequency; developing metric to consider sensitivity level of each PII attribute and combined sensitivity of a given set of leaked PII attributes; developed a novel adversarial adaptive defense (AAD) framework based on adaptive training; investigated interval-valued labels to enable a worker to specify both type-1 and type-2 uncertainties in his/her label without information loss; developed coded hate speech

detection framework, CODE, to judge coded words used in the coded meaning; performed exploratory study on fairness-aware design decision-making; conducted link prediction in identity network based on social network, intra-layer and inter-layer link information; documented and disseminated findings of literature research of privacy-preserving data analytics algorithms and software; conducted survey of existing work using cryptography for privacy protected in federated learning and designed a cryptography-based solution.

As recommended by the RSV this research group is revising their milestones and exploring more crossover with the Social Media team. The research team will be submitting a revised strategic plan for Years 4 and 5 in the coming weeks.

The Social Media and Networks component has lost some key personnel and is working to update the strategic plan and work more closely with Social Awareness as recommended by the RSV.

This component has determined key features and prepared the software design document for the cyber social network platform; developed, tested, and deployed data collection framework with real-time dashboard to monitor progress with alerting capabilities; revised taxonomy to characterize OIE based on social media platforms; studied cyber campaigns and characteristics of platforms and involved information actors and obtained data from Hurricane Harvey to represented a large-scale and geographically widespread disaster scenario.

Component researchers have produced 27 publications, 34 presentations and been awarded over $11 million dollars to advance their research. These publications include a taxonomy to characterize social media; various cyber campaign datasets; a model to identify contextual focal structures fundamental to coordinating cyber campaigns; one publication on water depth estimation and one on speed estimation based on social media data and routing algorithms used for single and multiple agent and destination variations of an uncertain and disrupted network. The progress in developing models and algorithms that can use Social Media postings to accurately inform disaster response is especially noteworthy.

The progress in SM1: Mining cyber argumentation data for collective opinions and their evolution and SM3: Auto-annotation of multimedia data are harder to gauge. The team reports the loss of key personnel and "human resource issues for graduate students" as partly to blame for lack of progress in these two areas. Perhaps revising the milestones and/or combining efforts with the Social Awareness team may be appropriate.

The Learning and Prediction research component is meeting most of its objectives.

This component has transitioned MTPP and LTSM to a convolutional neural network (CNN) approach; curated dataset of sensor data, system attributes, and failure/repair data of over 8,000 oil/gas wells; investigated efficacy of group structure on generalized neural network (GNN) architecture with smallest finite simple nonabelian group A5 action of random and clustered synthetic small size data; introduced new Deep learning algorithms that perform well in low-cost platforms with high accuracy which significantly reduce computational time and memory consumption; developed autoencoder method invention in unsupervised and self-supervised deep learning methods contributed to tackle problem of large-scale dataset management and labeling in Big Data management.

Developed a novel self-supervised domain adaptation method in crowd counting and published two novel tools-Fairness Domain Adaptation (FREDOM) and Approach to Semantic Scene Understanding and self-supervised Spatiotemporal Transformers (SPARTAN) available on GitHub. The group has also published research papers in IEEE ACCESS detailing novel features for classifying malware using machine learning models as well as published an image preprocessing model at the Intl. Conference on Image processing Computer Vision & Pattern Recognition. Developed and published a # of novel and best-in-class self-supervised tools for a variety of application: 3D capsule networks for medical segmentation on less labeled data; multimodality multi-lead ECG arrhythmias classification; image deblurring; self-supervised domain adaptation deep learning method to deal with limited training data. Published a model for CLIP-assisted temporal self-attention for weakly supervised anomaly detection.

The Education research theme efforts in postsecondary education are moving forward.

There are now three bachelor of data science programs in the state-UA at Fayetteville, A-State and UCA. The first cohort of graduates matriculated from UA at Fayetteville in May 2023. Five workshops on Data Science for Arkansas have been held during the first three years involving post-secondary institutions, ADHE and ACDS.

It is important to continue working with the cohort 1 institutions and begin work with cohort 2 institutions. The team should consider the recommendations by the EAB regarding courses to include in the Data Science degree programs in the state. An analysis of the historical data from ADHE on prior data science programs in the state may prove helpful in identifying those courses that have benefited graduates.

The Workforce Development component is working closely with the EAST Initiative to embed Data Science into its programs.

The Initiative is a 501-3C non-profit incorporated in 2006 and relies on a combination of legislative funds and corporate sponsors. Their mission is to provide K-12 students and teachers in Arkansas with access to technology and professional development to effectively integrate that technology into the curriculum. Partnering with EAST is a low cost and sustainable avenue for infusing data science into the educational fabric of public schools.

The Communication and Dissemination component has transitioned well from virtual meetings to face-to-face.

Both the annual retreats and annual conference are well attended and well received by participants. These events allow for networking among and between the research teams. The project website is used for communicating DART research to the broader research community and the public. The EAB recommended that the website be used to share research findings and datasets with other researchers.

The Broadening Participation component is achieving many of its goals by providing professional development for undergraduate and graduate students who tend to be female and URM.

The group has also awarded 2 broadening participation mini-grants of $5k each in Year 1 and 7 mini-grants a total of $15.5k in Year 2; hosted the ASRI for two years; supported undergraduates and graduate students each academic year and supported URM students each year in the SURE program.

The evaluation of the ASRI found that this program was engaging, educational, inspirational, comprehensive, rewarding and fun for the participants. Faced with the challenge of Covid-19 the leaders did a lot of planning and hard work to pivot from a face-to-face Institute to one that was totally virtual. They did not try to replicate the face-to-face training but instead thoughtfully designed a wholly new program that maximized the strength of a variety of virtual tools available to them and the students. While the ASRI could have been held face-to-face in Year 3, it was decided to continue the virtual program as it allows more student participants as well as more engagement by faculty.

This group needs the support of the research themes to achieve its diversity goals for: Faculty (45% female, 10% URM); Graduate Students (50% female, 20% URM) and Undergraduate Students (50% female, 40% URM). According to the self-reported participant data for Year 3 the project shows: Faculty (31% female; 4% URM); Graduate students (34% female; 12% URM) and Undergraduate students (45% female; 46% URM).

The Reverse Site Visit (RSV) panel met virtually because of COVID with the DART project team over zoom video/audio session on April 11-12, 2022. The DART management team presented their accomplishments and progress virtually to the panel.

While the RSV identified many areas of strength for DART, they also noted some areas for improvement. The panel's overall recommendations are listed below.

Recommendation 1. Website – Provide a plan for developing website sections listing the DART project's research publications and offering the project's datasets to the research community at large.

Recommendation 2. Cyberinfrastructure – (i) The panel did not find satisfactory details describing the architecture of the Arkansas Research Platform. Provide a detailed diagram showing the participating organizations, the connections, and the physical cyberinfrastructure. (ii) A significant sociotechnical challenge is federated identity management across participating institutions. Provide a workgroup document outlining the detailed needs assessment and the solutions for the sub-challenges identified for federated identity management. (iii) Provide a process document for the use of cyberinfrastructure during allocations, onboarding, and supporting user data operation needs. This can be coupled to an analysis of the state-scale common application themes based on interests, strengths, and needs. (iv) Implement a cyberinfrastructure risk register, with mitigation measures, and a monitoring dashboard to track and visualize cyberinfrastructure usage by various user groups.

Recommendation 3. Project management – Explain how the technical aspects of the DART project are being managed using the methods of formal project management, and who is qualified and responsible to conduct the above level of management.

Recommendation 4. Assessment – Explain how the evaluation plan provided by the project's external evaluator analyzes how research efforts, educational outcomes, seed grants, and student awards contribute to the project goals of enhanced research capacity.

The EAB has met two times during the project, once briefly during the first year virtually and a second time during the Year 2 DART Annual Meeting in Little Rock. Excerpts from the second report are presented below.

1. Coordinated Cyberinfrastructure

The Coordinated Cyberinfrastructure Team has a set of challenging aspects to address. First and foremost is the federated identity and access management of the respective campus research computing resources within the Arkansas Research Platform contributors.

Parallel to the federated access POCs, the EAB suggests the team explore methods to standardize the environment for collecting metrics from the ARP clusters. As ARP is further formalized and leveraging the respective computing clusters, collecting and reporting metrics across the participants becomes a priority to demonstrate the value. Open XDMoD would be a natural fit.

The EAB suggests the campus CISOs or designees of the participating campuses have a series of focused conversations and documentation of the practices and policies in place. In addition, they should include discussions with other regional campuses to compare existing multi-institutional research computing sharing practices. EAB suggests leveraging the SHARP-CI team.

2. Data Life Cycle and Curation

The EAB suggests that the team share their findings with the DART team and the broad scientific community. The idea of holding a team retreat this fall to revise the DC strategic plan is very good. It has the potential to bring the whole team on the same page in terms of language and terminology, to make a cohesive goal for this component, and the adoption of the proposed technologies.

Since facilitating the collaboration among multiple campuses in Arkansas is one of the goals of this project, the EAB suggests exchanging Ph.D. students among different campuses during the summer might be a good idea. Also, during the retreat or annual conference, bringing diverse groups together to discuss the proposal submission plan for the next year might be a good direction too. The EAB is delighted to see that the industry collaborations in this team are strong and continue to grow.

3. Social Awareness

The social awareness (SA) team has done a great job writing literature review papers and survey papers on privacy, statistical fairness, and causal fairness. They have developed various algorithms to provide privacy preservation, fairness, safety, and robustness of data analytics, data collection, data sharing, and decision making, with multiple publications in leading conferences and multiple awarded grants and training of numerous graduate students. All seven projects have been conducted as scheduled. It seems that SA1 and SA2 have been completed. The EAB would like to see the SA team plans to expand the activities, have better communication of SA work and results, and invite a social scientist member to the team.

4. Social Media & Networks

The EAB believes that good progress has been made and that this component is meeting most of its objectives. The SM team has determined key features and software design document for the cyber social network platform; developed, tested, and deployed a data collection framework with real-time dashboard to monitor progress with alerting capabilities; revised the taxonomy to characterize OIE based on social media platforms; studied cyber campaigns and characteristics of platforms and involved information actors and selected Hurricane Harvey to represent a large-scale and geographically widespread disaster scenario. Research into the auto-annotation of multimedia data goal appears to be trailing the progress being made in the other goals/objectives, as noted by the external reviewers. Please address the concerns raised by the evaluator, including "obtain and index content types for at least two disaster scenarios" and "developing GIS system to display real-time road status inputs" have not been done yet.

The EAB suggests the project team consider the following questions raised in the RSV report as potential areas of interest for the next step, including: (i) how to incentivize users to provide trusted information; (ii) how to develop trust models and handle biases; and (iii) how to reliably assess information-driven context or situation awareness leading to safety and security. The number of papers generated in this component is beyond expectation; however, the EAB suggests that the team keep improving the quality of the work by publishing in top conferences in the field. Targeting the conferences listed on ankings.org might be a good start.
Also, as one of the co-leaders of the team left the university, the EAB suggests that the project leadership take action as soon as possible to revise the SM strategic plan in consideration of the new team member's expertise.

5. Learning and Prediction

The EAB noted that the team implemented the EAB suggestions from last year with regard to greater elaboration on its achievements. The EAB suggests connecting Learning and Prediction more closely with other project areas, such as Social Awareness. It is indeed exciting to note that this component has grown to cover a wide range of important activities that are core to the DART mission. Therefore, the EAB suggests that the team consider structured/facilitated ideation workshops involving the different LP project members with project members from other components (e.g., Data Curation, Social Awareness, Social Media and Networks, Education) to facilitate cross-fertilization of ideas and possibly address the RSV concern of impact and outreach both within DART and its industry partners and across the broader community.

6. Education

The EAB noted that the DART educational team reports significant postsecondary initiatives across six campuses spanning both community colleges and universities within the state. Additionally, some key efforts in the K-12 space were also noted. Finally, the DART ED team reports several strong government, industry, startup (including through the NSF I-Corps program) and non-profit collaborations which can help improve educational outcomes and opportunities for graduates.

The EAB reiterates its suggestion from last year to emphasize the importance of continuous evaluation of programmatic elements – this becomes increasingly important as the program's first graduates join the workforce next year. The EAB was pleased to see examples of ethics in data science curriculum development and would like to see more details on this element in next year's report.

The EAB does have several suggestions for the DART team to consider. First –given the impressive array of collaborations identified – board members suggest leveraging these industry and government collaborations to help amplify and highlight the importance of data science careers in local industries. Second, such connections can be leveraged in capstone projects within the curricular infrastructure being proposed. Third, research projects from other components of DART (e.g., seed grant recipients) can serve be leveraged in such courses and can serve as exemplars of research in pedagogy. Finally, as the first graduates come through the program, the DART ED team needs to start to think about holistic programmatic outcome assessments.

7. **Workforce Development and Broadening Participation**

One possible way to attract more attention to the SURE program could be to reach out to the smaller universities in Arkansas, for instance, the PUIs, and the HBCUs. These are more likely to have students with less access to research opportunities in areas covered by the DART project. Thus, working with faculty at these smaller institutions/universities and then reaching out to their students could improve awareness of the SURE opportunity for such students. Also, providing the SURE program as a summer internship for the target students might help enhance interest in the program.

To encourage faculty in the smaller universities to be more involved with the SURE program, DART could also provide incentives for faculty members that take up SURE students, for instance, by providing summer support to the faculty. Since most of these faculty members are involved in teaching to cover their summer, this could also help them put more effort into their own research activities, especially during the summer.

8. **Communication and Dissemination**

Communication and dissemination are extremely important, and maintaining internal communication across campuses and research is challenging. The EAB provides a variety of possible technology stacks that have been used in different campuses successfully. The two main ecosystems are Microsoft Teams and Open Source.

MS Teams can serve as a highly integrated platform for chat, conference calls and file transfer. The main challenges people have faced are permissions and user management. If all campuses in DART already have Microsoft Teams subscriptions, this may be a more or less complicated option. One problem for MS Teams may be centralized user management based on AD. This may simplify the setup here if this is set up at the central office level. MS Teams is not as powerful for real-time communications, which may be a desirable feature for close collaboration between DART members and groups.

An alternative tech stack is based on separate, freely available components. The main piece that we have seen being used is a discord for real-time communications, which is already used by many contributors already and makes it fairly simple to add a DART-specific group. For longer-term communications, we have seen good success with using web fora like phpBB or similar systems for recorded conversations. For maintaining shared definitions or onboarding documents, a shared document platform like a wiki like MediaWiki (or many others) has proven useful. Other shared document options are possible, but Wikis seem common enough for people to accept them fairly easily and successfully. This will need some seeding, possibly from the central coordination group for DART, but once the basics are provided, Wikis tend to grow well.

For maintaining more project progress-related communications, workflow tools like Trello or Asana can easily maintain tasks and progress histories in an easy-to-understand and digest format. When used consistently, they can significantly simplify new user onboarding. Neither is Open Source, but they have free usage tiers for smaller projects. There are Open Source alternatives like taiga.io and others if needed.

The EAB believes that different technological solutions will enable effective post-covid communications and would encourage the project team to pick the easiest to integrate and closest to what is being used in the different groups already. We have had limited success with adding too many new components to existing environments unless no tech stack is overlapped between the admittedly pretty heterogeneous and fairly large number of groups in DART. We would encourage the central group to survey tools already in use and find the intersection between all the DART groups. For dissemination, the project website is the primary tool, and it already does so very well. Including a section that clearly highlights project outcomes would be beneficial.

The Management, Evaluation, and Assessment Component through its project leadership has worked to ensure that the research teams have benefited from the feedback provided by both the EAB and RSV panel.

The leadership provided time at the annual retreat for the external evaluator to present the findings from these two external reviews and to encourage the research groups to revise their strategic plans based upon these findings. The IAB has been less involved in providing feedback but has been involved in sharing industry needs with the DART project.

DART has more than tripled the number of research groups that need to be monitored with the awarding of 16 seed grants. In addition, the project has also awarded broadening participation seed grants to several institutions. The leadership team may want to consider expanding the central office staff to include another professional to help oversee these additional project activities and ensure that these awards fulfill their promises and contribute to the overall mission of DART.

Overall, DART has ‘Met/Exceeded Targets’ in 30% of its strategic plan metrics, are ‘On Track’ to meet in 52% and are ‘Delayed’ in 18% of the strategic plan metrics.

The Cyberinfrastructure Component has ’Met/Exceeded Targets’ in 73% of its metrics, are ‘On Track’ to meet in 9% and are ‘Delayed’ in 18%. Those metrics which are delayed are: ‘Instructors trained in software carpentry’ and the number of ‘Online workshops’.

The Research Component has ‘Met/Exceeded Targets’ in 67% of their metrics, are ‘On Track’ to meet in 22% and are ‘Delayed’ in 11%. The metric which is delayed is the proposed number of ‘Applications and Platforms’ in the various research thrusts.

The Education and Workforce Development Component has ‘Met/Exceeded Targets’ in 10% of their metrics, are ‘On Track’ to meet in 71% and are ‘Delayed’ in 19%. The metrics which are delayed are: number of ‘Seed grants awarded to K-12 teachers’, ‘number of industry internships awards’, ‘% internship students supported who are URM/female’ and the development of ‘Industry internship evaluation form’.

The Communication Component ‘Met/Exceeded Targets’ in 30% of their metrics, are ‘On Track’ to meet in 60% and ‘Delayed’ in 10%. The metric which is delayed is the ‘# of DART topical webinars held’.

The Broadening Participation component has ‘Met/Exceeded Targets’ in 6% of its metric, are ‘On Track’ in 67% and ‘Delayed’ in 27%. Those metrics which are delayed are: ‘Provide training to mentors on how to use IDP and to be a mentor’, ‘Early career faculty complete IDP templates pre and post’, ‘Award maximum of $8k per SURE faculty’ and ‘SURE Students complete IDP template at beginning and end of experience’.

APPENDIX

Appendix A
Strategic Plan Metrics: DART Wide Research Thrusts
Strategies, Outputs, Targets, Progress and Status

Strategies	Outputs	Targets	Year 1	Year 2	Year 3	Current Status
Computer Infrastructure						
Hardware and software infrastructure improvements	Install, configure, and make available data science nodes on Pinnacle Portal	1	0	1	0	Met Target
	Science DMZ at UA and UAMS/UALR	1	0	1	0	Met Target
	100GB connection between ScienceDMZs	1	0	1	0	Met Target
	Establish dedicated DART Gitlab ~~repository~~ organization	1	0	0	1	Met Target
	Setup Globus data management services to point at DART storage arrays	1	1	0	0	Met Target
Documentations and user guides	Create technical management document defining organizational structure, roles, and responsibilities of ARCC	1	1	0	0	Met Target
	Amend existing MOU for ARCC expansion	1	0	1	0	Met Target
	UAF and UAMS will create CI Plans to support DART (1 x UAMS)	2	1	1	0	Met Target
	Create and publish document outlining GitLab user guidelines and minimum standard for code repository	1	0	0	0	On Track
Instructors trained in software carpentry	CI=8	8	0	2	2	Delayed (4)
Workshops (Online)	CI=30; DC=0; SA=0; SM=0; LP=0	30	NA	5	5	Delayed (10)
Research Themes						
Journal publications	CI=0; DC=28; SA=7; SM=9; LP=10	54	8	89	46	Exceeded Target (143)
Develop software modules/prototypes	CI=0; DC=0; SA=0; SM=3; LP=0	3	0	1	1	On Track (2)
Conference presentations, seminars, papers and posters	CI=0; DC=25; SA=16; SM=5; LP=14	60	42	45	56	Exceeded Target (143)
Invention disclosures/patents	CI=0; DC=2; SA=0; SM=0; LP=0	2	0	0	6	Exceeded Target (6)
Data Sets and algorithms/mathematical formulation	CI=0; DC=20; SA=4; SM=12; LP=1	37	2	5	28	On Track (35)
Applications and Platforms	CI=5; DC=0; SA=0; SM=2; LP=0; WD=1	8	0	1	1	Delayed (2)
Workshops, demonstrations/trainings	CI=4; DC=5; SA=0; SM=0; LP=3; ED=16; WD=5	33	15	18	2	Exceeded Target (35)
Training and Webinars (Online)	CI=0; DC=0; SA=0; SM=0; LP=0; WD=8	8	1	5	2	Met Target (8)
Proposals Submitted	CI=0; DC=0; SA=7; SM=0; LP=0/1	8	35	42	19	Exceeded Target (96)

Appendix B

Education and Workforce Development

Strategies, Outputs, Targets, Progress and Status

Strategy	Outputs	Target	Year 1	Year 2	Year 3	Current Status
Education & Workforce Development						
~~Middle school curriculum developed~~	~~Posted at ADE and Co-ops~~	~~12~~	~~0~~	~~2~~	~~Removed~~	~~Removed~~
	~~Piloted in schools~~	~~6~~	~~0~~	~~1~~	~~Removed~~	~~Removed~~
Undergraduate Data Science degree program	Certification by ADE	1	0	0	0	On Track
	Evaluation report	2	0	0	1	On Track
	Implemented on campuses	3	1	1	2	Exceeded Target (4)
~~Collaborate with DART Faculty to develop capstone projects~~	~~Develop capstone projects~~	~~25~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
	~~Publish capstone projects~~	~~15~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
	~~Student presentations at AHM~~	~~4~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
Collaborate with ACDS to co-host workshops on data science topics	# held/# attending	5/100	1/20	1/20	2/40	On Track (4/80)
Annual workshops for faculty on grantsmanship and entrepreneurship topics	# held/# attended	15/100	1/20	1/20	2/57	On Track (4/97)
Seed grants for K-12 teachers	Number awarded	50	0	7	1	Delayed (8)
	~~Awardee presentations at AHM~~	~~15~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
~~Training grants for higher ed faculty~~	~~Number awarded~~	~~50~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
Present (booth or breakout) at annual EAST conference	Number completed	5	1	1	1	On Track (3)
Training Webinars	Number held	8	1	1	5	On Track (7)
Training sessions for K-12 teachers	Number held/# teachers attending	4/200	1/20	2/49	3/52	On Track (5/121)
Sharing platform for K-12 teachers	Established/# using	1/50	NA	NA	1/NA	On Track
Undergraduate research assistantships during school year	# assistantships/# unique students	125/75	25	41	51	On Track (117/62)
	% students supported URM/female	35%/25%	36%/48%	45%/46%	46%/45%	On Track (46%/48%)
Undergraduate summer research experience	# URE/# unique students	50/35	10/10	10/10	15/15	On Track (35/35)
	% students supported URM/female	50%/50%	NA	26%/51%	53%/20%	On Track (53%/20%)
Graduate research assistants	# assistantships/#students	200/100	47	87	116	Exceeded Target (250/118)
	% students supported URM/female	50%/50%	13%/38%	12%/30%	12%/34%	On Track (13%/35%)
Industry internships	# internships/# students	20/10	NA	NA	3	Delayed (3/3)
	% students supported URM/female	25%/35%	NA	NA	0%/33%	Delayed (0%/33%)
	Evaluation form	1	NA	NA	NA	Delayed
Master Theses	CI=0; DC=0; SA=7; SM=0; LP=3	10	1	NA	7	On Track (8)
PhD Dissertations	CI=0; DC=7; SA=0; SM=0; LP=2	9	1	NA	7	On Track (8)

Appendix C

Communication Component

Strategies, Outputs, Targets, Progress and Status

Strategy	Outputs	Target	Year 1	Year 2	Year 3	Current Status
Communications						
All Hands Meeting	# held/# attending per meeting	5/100	1/110 (virtual)	1/125	1/123	On Track 3/358
	~~K-12 teacher presentations~~	~~4~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
	Poster competitions/# participants per competition	5/25	1/10	1/20	1/50	On Track (3/80)
Annual retreat for faculty and grad students	# held/# attending per retreat	5/25	Covid	1/125 (virtual)	1/58	On Track (2/283)
Monthly DART topical webinars	#held	55	5	6	7	Delayed (18)
Monthly DART component team meetings	#held	330	88	96	53	On Track (237)
Project website	Published	1	0	1	NA	Met Target
Quarterly Blogs posts about project	# posts	20	4	5	2	On Track (11)
Maintain Facebook, Twitter and YouTube channels	Increase Following 10% a year	50%	7%	50%	NA	Met Target (50%)
ER Core reporting system	Published and accessible	1	1	NA	NA	Met Target
	Annual webinar training	15	3	3	3	On Track (12)
~~Science Journalism Challenge~~	~~Host/# attending per year~~	~~4/20~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~
	~~Invite SJC winners to AHM~~	~~20~~	~~NA~~	~~NA~~	~~Removed~~	~~Removed~~

Appendix D
Broadening Participation
Strategies, Outputs, Targets, Progress and Status

Strategy	Outputs	Target	Year 1	Year 2	Year 3	Current Status
Mentorship of students and early career faculty	Updated versions of University of Hawaii IDP templates	3	NA	3	NA	Met Target (3)
	Provide training to mentors on how to use IDP and to be a mentor	50	NA	NA	NA	Delayed
Research Seed Grants Program	Develop and widely distribute RFP for research seed grants	5	1	1	1	On Track
	In conjunction with EAB and IAB select best RFPs and award funding	50	2	16	6	On Track (24)
	Research seed grant presentations to DART community	25	1	2	6	On Track (9)
	Early career faculty complete IDP templates pre and post	25	NA	NA	NA	Delayed
	Connect early career faculty with senior faculty	25	2	16	6	On Track (24)
Summer Undergraduate Research Experienced (SURE) Program	Develop and widely distribute SURE RFP for DART faculty to host students	50	10	10	10	On Track (30)
	Award maximum of $8k per faculty award	25	NA	NA	NA	Delayed
	Recruit URM students	50	10	10	10	On Track (30)
	Students complete IDP template at beginning and end of experience	50	NA	0	NA	Delayed
Host ASRI	# held/# attending per ASRI	5/35	NA	1/42	1	On Track (2/##)
	% students supported URM/female	50%/50%	NA	26%/51%		On Track
	Evaluation report	4	NA	1	1	On Track
	Scholarships for underserved	100	NA	Virtual	Virtual	On Track