ADSA Spring Meeting: Print Schedule

[B Breakout](#)

[K Keynote](#)

[P Panel and/or Discussion](#)

March 7 • Monday 8:30am – 9:00am [Welcome, coffee and pastries](#)
Atrium

Speakers: Phil Bourne, Micaela Parker
Coffee and pastries in the atrium. Welcome remarks and some logistics.

9:00am – 10:45am P [Data Science Program Sustainability](#)
Atrium

Speakers: Phil Bourne, Jay Aikat
This session explores the organizational journey from the startup phase to sustainability that every DS program, in universities across the research and teaching spectrum, needs to consider. Different institutions have different priorities and strategies to achieve their goals, and this often depends on how their research capacity intersects with the teaching mandates of their departments. While acknowledging these differences, we wish to explore and understand commonalities in these organizational journeys specific to the creation and building of sustainable data science initiatives.
Chairs: **Jay Aikat** (University of North Carolina, Chapel Hill) and **Phil Bourne** (University of Virginia)

10:45am – 11:00am [Break, Setup Posters](#)
Main Hallway

11:00am – 12:00pm P [University Update Lightning Posters](#)
Atrium

Moderators: Phil Bourne, Amy Wagler, Jay Aikat
Speakers: LaTanya Brown-Robertson, Jill Keith, Talitha Washington, David Bader, Jing Liu, Sarah Stone, Jason T. Black, Azene Zenebe, David Mongeau, H.V. Jagadish, Rachel Saidi, Azer Bestavros, David Ebert, Norene Kemp, Andrew Connolly, Debzani Deb

The goal of these sessions is to provide an opportunity for heads of academic data science programs to share their challenges, how they overcame these challenges, and provide opportunities for sharing of lessons learned and solutions along the way.

Read more about participating institutions here.

Part 1: Successes, Challenges, and Best Practices for Data Science Programs at HBCUs
Chairs: **LaTanya Brown-Robertson** (Bowie State) and **Talitha Washington** (Atlanta University Center Consortium, Data Science Initiatives

Speakers
:
Jason Black, Florida Agricultural and Mechanical University
Sajid Hussain, Fisk University
Jill Keith and Debzani Deb, Winston-Salem State University
Venkata Melapu, Jackson State University
Felisia Stukes, Johnson C Smith University
Talitha Washington, Atlanta University Center Data Science Initiative
Azene Zenebe, Bowie State University

Part 2: Successes, Challenges, and Best Practices for Data Science Institutes and Programs Nationwide
Chairs: **Andreas Berlind** (Vanderbilt University) and **Amy Wagler** (University of Texas, El Paso)

Speakers
:
Coming soon!

12:00pm – 1:30pm Lunch, Poetry, and Fresh Air break
Dining room/Courtyard

1:30pm – 2:20pm K Partnerships with Industry Keynote
Auditorium

Speakers: Claire Le Gal
Claire Le Gal
Senior Vice President, Fraud Intelligence, Strategy and Cyber Products

Claire Le Gal currently is Senior Vice President, responsible for the Fraud Intelligence, Strategy and Cyber Products team in the Cyber & Intelligence Solutions division at Mastercard. In that role, she is responsible for a variety of global functions, including fraud reporting & analytics, cyber security

solution development, and the execution of strategic initiatives to ensure that the integrity of the MasterCard payment systems is maintained as payment technologies evolve. Claire also chairs the MasterCard Global Fraud Advisory Council and the Mastercard Global Payment Integrity Coalition.

2:20pm – 2:30pm [Break](#)
Atrium

2:30pm – 3:30pm P [Partnerships with Industry Panel Discussion](#)
Auditorium

Moderators: Sharon Sputz, Oscar Wood
Speakers: Karina Edmonds, Kory Hawkins, David Mongeau
The number of Data Science programs across academia continues to increase to match industries' need for talent. With this comes the challenge we are all facing to improve our partnerships with industry. These partnerships serve many purposes: to raise funds, to provide access to data, to enable project-based learning for our students, and to drive research for impact. In this session, we will have representatives from industry and academics who have had a long history of success in running Academic-Industry programs. We will explore different models and have engaging open conversations on challenges and solutions.
Chairs: **Sharon Sputz** (Columbia University) and **Oscar Wood** (NNData Corporation)

Confirmed Panelists
Karina Edmonds, SVP, Global Head of Academies and University Alliances, SAP
Kory Hawkins, University Relations Lead, Microsoft
David Mongeau, University of Texas, San Antonio

3:30pm – 4:30pm B [Partnerships with Industry Breakout Groups](#)
Atrium

Speakers: Sharon Sputz, Erwin Gianchandani, Oscar Wood 4:30pm – 4:50pm B [Partnerships with Industry Report Out](#)
Auditorium

5:00pm – 6:00pm [Research Posters, Art installation, Networking Happy Hour](#)
Main Hallway and Courtyard

Speakers: Sunandan Chakraborty, Hawley Helmbrecht, Daniel Katz, Jacqueline Wernimont
Energy Pools: Data Visceralizaton of Clean Energy Transitions
Jacqueline Wernimont (Dartmouth College) and Nikki Stevens (Dartmouth College/Arizona State

University)

Energy Pools is an interactive and tactile installation created by in a faculty-study collaboration to help people understand energy usage in the United States. Climate and data scientists continue to search for ways to make large scale data sets and long-term problems like those associated with anthropogenic climate change accessible to non-specialist audiences. Energy Pools visceralizes data - rendering it in three dimensions and encouraging embodied interaction - rather than only visualizing the data. We do this in order to communicate U.S. energy consumption by source and to help people understand the scale of change that is needed to keep global warming at 1.5 degree Celsius by 2050. Created from recycled textiles sourced from women-owned small businesses, the piece also uses its material form to argue that we can produce cutting edge knowledge using traditionally feminized forms of production.

Non-destructive Microglia Phenotype Quantification with Image Processing in Immunofluorescent Brain Images

Hawley Helmbrecht (University of Washington) and Elizabeth Nance (University of Washington) Microglia are the resident immune cells of the brain and can change morphology and function in response to environmental stimuli and injurious events. Therefore, microglia have become of interest for therapeutic targets. However, microglia are dynamic and when analyzed on a population-level, exhibit heterogenous phenotypes that complicate therapeutic strategies. We sought to provide a non-destructive, quantitative imaging-based assessment to categorize individual and population level microglial features based on microglia shapes. We aimed to quantify phenotype heterogeneity across treatments, brain regions, sex, and species. We partnered image-processing powered geometric analysis of individual, segmented microglia with principal component analysis of large regional populations of microglia in neonatal rats and ferrets. In rats, regional populations of microglial express higher phenotypic heterogeneity after ischemic injury compared to a non-treated control. Additionally, we applied geometric analysis to individual microglia to analyze circularity and observed increased circularity with injury that correlates with microglia-driven pathogenic neuroinflammation. When we applied our data science pipeline to the ferret, we observed microglia-response to injury that varied across brain region and sex. Population-wide sex differences were observed in microglia phenotype distributions. When populations were analyzed by region, we observed differences in the reactivity of microglia to injury across the hippocampus, cortex, and corpus callosum – this data aligns with transcriptional data about microglial roles across brain regions. By developing quantitative cell morphology pipelines for microglial images, we can quantify population-level microglia phenotype presentation and individual microglia features for comprehensive understanding of the dynamic behavior of microglia.

Exploring Dyanmism of Causal Relationships Extracted from Text

Sunandan Chakraborty, Jack VanSchaik, Palak Jain, Sawyer Lehman, Taofeek Alao (Indiana

University)

Text mining made important advances in methods to convert vast and unstructured text data into knowledge. However, the current paradigm of relationship extraction has one major limitation: it fails to capture the fundamentally dialogic and dynamic nature of knowledge: This NSF-funded project aims to capture such dynamics of knowledge, specifically focusing on causal relationships. The objective of this project is to identify cues of causal knowledge from text data, quantify the strength of the causal relationship, and model its dynamics over changing conditions. As text data is extensively used by researchers and practitioners from different domains of national importance, including, medicine and health, the results of this project seek to provide the foundation to offer practitioners new ways to understand the evolving nature of the causal relationships present in large text datasets.

We utilize the semantics of sentences depicting causality to identify the influence of certain words in the sentence to characterize the nature of the causal relationship expressed in the text. This task captures the strength of the relationship (e.g., "extremely likely"), any supporting or opposing evidences (e.g., "will lead to" or "does not lead to") to encode the "causal distance" between a relationship pair. Causal distance captures the magnitude of causality between two entities, while capturing the dynamism of the relationship by modifying itself over time with changing conditions or new evidences. We evaluate our methods and the utility of causal distance on a corpus of research articles taken from the PubMed database.

Portable serverless computing to enable scalable data science

Daniel S. Katz (University of Illinois, Urbana-Champagne) and Kyle Chard (University of Chicago)

Data science depends on data, algorithms/software, computing, and people. In the past, a data scientist would work directly on the computing resource where they wanted to run their algorithms (e.g., their laptop, a shared cluster, or the cloud). More recently, the data scientist could develop containers with their software and deploy them to the computing resources. Today, serverless computing is an even lighter-weight and easier-to-program option, where only functions are deployed from a development platform to an execution platform. While the architecture that controls the deployment and execution of these functions is often specific to a platform (e.g., Lambda for AWS, Cloud Functions for Google, etc.), we've developed a portable serverless computing platform called funcX (https://funcx.org) that allows a data scientist to portably run their functions on various cloud, local, or HPC resources, including campus clusters and NSF- and DOE-funded HPC systems. These functions can be run where resources are available soonest or cheapest, or where data exists (e.g., a cluster close to an instrument that produces data in real time). By using a cloud-hosted service, functions can be launched in a fire-and-forget style, and can be shared in a user-controllable manner. funcX has been used for problems such as scalable fitting and columnar analysis in High Energy Physics, generating training data for molecular dynamics machine learning, running ensembles of COVID-19 epidemiologic models, real-time experiment analysis at beamlines including Argonne's

Advanced Photon Source, scalably executing publicly-stored ML models, and text extraction.

March 8 • Tuesday 8:50am – 9:00am Announcements
Auditorium

9:00am – 10:00am K Impacted Communities Keynote
Auditorium

Speakers: Desmond Upton Patton
Keynote abstract coming soon!

Dr. Desmond Upton Patton, Associate Dean for Innovation and Academic Affairs, founding director of the SAFE Lab and co-director of the Justice, Equity and Technology lab at Columbia School of Social Work, is a leading pioneer in the field of making AI empathetic, culturally sensitive, and less biased. Through keynote presentations and interactive workshops, Patton is helping organizations develop a better approach to diversity and inclusion that includes fairer practices that address the challenge of prejudice, rather than contribute to it. Also the Associate Director of Diversity, Equity, and Inclusion and the co-chair of the Racial Equity Task Force at The Data Science Institute and founder of the SIM|ED tech incubator at Columbia University, Patton's research uses virtual reality to educate youth and policymakers about the ways social media can be used against them and how race plays a part.
In 2018, Professor Patton published a groundbreaking finding in the prestigious Nature journal, Digital Medicine, which uncovered grief as a pathway to aggressive communication on Twitter. The report was cited in an amici curiae brief submitted to the United States Supreme Court in Elonis v. United States, which examined the interpretation of threats on social media. Widely referenced across disciplines, Patton's research at the intersections of social media, AI, empathy and race has been mentioned in the New York Times, Nature, Washington Post, NPR, Vice News, ABC News and other prestigious media outlets more than seventy times in the last three years.

10:00am – 10:30am Break, coffee and pastries
Atrium

10:30am – 12:00pm P Impacted Communities Panel + Discussion
Auditorium

Moderators: Sharon Tettegah, William Wang
Speakers: Faithe Day, Tony Baylis, Francisco J. Marmolejo-Cossío, Safiya Umoja Noble
This session will present information on how data science affects impacted people of color and their communities. Various aspects of impacted communities will be discussed. For example, How is the

collection and use of data affecting these communities? What are the implications for biased data, and how does that data affect Black, Chicanx, Native communities informally (social spaces) and formally (academic spaces)? What are the implications for AI and ML within and between communities of color and women? How can we engage in data justice for all? How can academia set norms that are anti-racist and inclusive for all communities? We will discuss the fundamental challenges and opportunities centering this topic, and in the session, we will also collectively brainstorm responses to important questions that improve the diversity, equity, and inclusiveness of our data science community.

Chairs: **Sharon Tettegah** (University of California, Santa Barbara) and **William Wang** (University of California, Santa Barbara)

Panelists
:

Tony Baylis at Lawrence Livermore National Laboratory is the Laboratory's Director for the Office of Strategic Diversity and Inclusion Programs. In this position, he is the senior management advocate for diversity and inclusion for the Laboratory.

Tony is responsible for overseeing the laboratory's interactions and successful execution in building, partnering and collaborating with governmental, educational, industrial, community interests and other stakeholders. LLNL has had a long history in working with Minority Serving Institutions, specifically relationships with American Indian Institutions, Hispanic Institutions and Historically Black College and Universities. He represents the Laboratory on the subjects of Diversity and Inclusion, STEM, Outreach Efforts, and Student Programs.

Tony's career represents 29 years of administrative, project, program, technical and organizational management. He has worked in a scientific and technical environment for over 20 years and has worked as an consultant in industry as well. Tony has extensive experience networking with a broad range of academic, industry, government and non-profit organizations that has educated him and helped him in his career. He is a DOE Minorities in Energy Champion for the department and also serves on a number of conference program committees and advisory boards that promote STEM and diversity in science and technical careers.

Faithe J. Day is a Writer, Creator, and Educator with a B.A. in English and Digital Humanities, as well as a PhD in Communication Studies. Currently, Dr. Day is a Postdoctoral Fellow at the Center for Black Studies Research working to imagine a more just future through collaboration with mission focused organizations, individuals, and initiatives, such as the Student Enrichment for Engagement in Data Science (SEEDS) Program. Through building a culture of collaboration within digital platforms and online communities, Dr. Day develops creative solutions to 21st century problems, all while educating and empowering intersectionally marginalized communities. Dr. Day is also the creator of the "Black Living Data Booklet" a manual and a manifesto on the ethical engagement of data on and

for Black and LGBTQIA+ communities. Her work brings greater awareness to issues of data privacy, ethics, and ownership in developing personal and community based archives.

Francisco Marmolejo-Cossío is a Postdoctoral Researcher at the Harvard School of Engineering and Applied Sciences (SEAS) and a Research Fellow at Input Output Hong Kong (IOHK). Prior to this, he was a Career Development Fellow in Computer Science at Balliol College at the University of Oxford. He completed a D.Phil. in Theoretical Computer Science under the supervision of Paul Goldberg, and a B.A. in Mathematics at Harvard University with a minor in Neuroscience in 2012. His academic interests lie in the intersection of Algorithmic Game Theory, Decentralised Consensus Protocols, and Computational Learning Theory. He is also a co-organiser of the Mechanism Design For Social Good (MD4SG) research initiative.

Safiya Umoja Noble is an associate professor at UCLA, and is the Co-Founder and Co-Director of the UCLA Center for Critical Internet Inquiry.[1] She is the author of *Algorithms of Oppression*, and co-editor of two edited volumes: *The Intersectional Internet: Race, Sex, Class and Culture* and *Emotions, Technology & Design.* She is a Research Associate at the Oxford Internet Institute at the University of Oxford. She was appointed a Commissioner to the University of Oxford Commission on AI and Good Governance in 2020.[2] In 2020 she was nominated to the Global Future Council on Artificial Intelligence for Humanity at the World Economic Foundation.[3]

12:00pm – 2:00pm Group photo, then Lunch, Poetry, Lawn games!
Dining room/Courtyard

Speakers: David Mongeau, Amy Wagler 2:00pm – 3:00pm P Engaging Broadly
Auditorium

Moderators: David Ebert, Miriam Posner
Speakers: Katharine Coles, Stuart Geiger, Gillian R. Hayes, Erin Maher
This session will explore and discuss the bidirectional opportunities of data science with non-STEM fields, as well as best practices of engaging researchers and practitioners from fields that have invested in and studied the social process of the construction of data. What can we learn from cultural analytics, political science, sociology, philosophy, history, the digital humanities, and communication theory to help humans more effectively use data science, as well as help researchers understand how social factors might influence the way their data itself was constructed?

3:00pm – 4:00pm B Engaging Broadly, Breakout Tables
Atrium

Moderators: David Ebert, Miriam Posner
Speakers: Katharine Coles, Stuart Geiger, Gillian R. Hayes, Erin Maher
This session will explore and discuss the bidirectional opportunities of data science with non-STEM fields, as well as best practices of engaging researchers and practitioners from fields that have invested in and studied the social process of the construction of data. What can we learn from cultural analytics, political science, sociology, philosophy, history, the digital humanities, and communication theory to help humans more effectively use data science, as well as help researchers understand how social factors might influence the way their data itself was constructed?

4:00pm – 5:30pm Research Posters, Art installation, Networking Happy Hour
Main Hallway and Courtyard

Speakers: Sunandan Chakraborty, Hawley Helmbrecht, Daniel Katz, Jacqueline Wernimont
Energy Pools: Data Visceralizaton of Clean Energy Transitions
Jacqueline Wernimont (Dartmouth College) and Nikki Stevens (Dartmouth College/Arizona State University)
Energy Pools is an interactive and tactile installation created by in a faculty-study collaboration to help people understand energy usage in the United States. Climate and data scientists continue to search for ways to make large scale data sets and long-term problems like those associated with anthropogenic climate change accessible to non-specialist audiences. Energy Pools visceralizes data - rendering it in three dimensions and encouraging embodied interaction - rather than only visualizing the data. We do this in order to communicate U.S. energy consumption by source and to help people understand the scale of change that is needed to keep global warming at 1.5 degree Celsius by 2050. Created from recycled textiles sourced from women-owned small businesses, the piece also uses its material form to argue that we can produce cutting edge knowledge using traditionally feminized forms of production.

Non-destructive Microglia Phenotype Quantification with Image Processing in Immunofluorescent Brain Images
Hawley Helmbrecht (University of Washington) and Elizabeth Nance (University of Washington)
Microglia are the resident immune cells of the brain and can change morphology and function in response to environmental stimuli and injurious events. Therefore, microglia have become of interest for therapeutic targets. However, microglia are dynamic and when analyzed on a population-level, exhibit heterogenous phenotypes that complicate therapeutic strategies. We sought to provide a non-destructive, quantitative imaging-based assessment to categorize individual and population level microglial features based on microglia shapes. We aimed to quantify phenotype heterogeneity across treatments, brain regions, sex, and species. We partnered image-processing powered geometric analysis of individual, segmented microglia with principal component analysis of large regional

populations of microglia in neonatal rats and ferrets. In rats, regional populations of microglial express higher phenotypic heterogeneity after ischemic injury compared to a non-treated control. Additionally, we applied geometric analysis to individual microglia to analyze circularity and observed increased circularity with injury that correlates with microglia-driven pathogenic neuroinflammation. When we applied our data science pipeline to the ferret, we observed microglia-response to injury that varied across brain region and sex. Population-wide sex differences were observed in microglia phenotype distributions. When populations were analyzed by region, we observed differences in the reactivity of microglia to injury across the hippocampus, cortex, and corpus callosum – this data aligns with transcriptional data about microglial roles across brain regions. By developing quantitative cell morphology pipelines for microglial images, we can quantify population-level microglia phenotype presentation and individual microglia features for comprehensive understanding of the dynamic behavior of microglia.

Exploring Dyanmism of Causal Relationships Extracted from Text

Sunandan Chakraborty, Jack VanSchaik, Palak Jain, Sawyer Lehman, Taofeek Alao (Indiana University)

Text mining made important advances in methods to convert vast and unstructured text data into knowledge. However, the current paradigm of relationship extraction has one major limitation: it fails to capture the fundamentally dialogic and dynamic nature of knowledge: This NSF-funded project aims to capture such dynamics of knowledge, specifically focusing on causal relationships. The objective of this project is to identify cues of causal knowledge from text data, quantify the strength of the causal relationship, and model its dynamics over changing conditions. As text data is extensively used by researchers and practitioners from different domains of national importance, including, medicine and health, the results of this project seek to provide the foundation to offer practitioners new ways to understand the evolving nature of the causal relationships present in large text datasets.

We utilize the semantics of sentences depicting causality to identify the influence of certain words in the sentence to characterize the nature of the causal relationship expressed in the text. This task captures the strength of the relationship (e.g., "extremely likely"), any supporting or opposing evidences (e.g., "will lead to" or "does not lead to") to encode the "causal distance" between a relationship pair. Causal distance captures the magnitude of causality between two entities, while capturing the dynamism of the relationship by modifying itself over time with changing conditions or new evidences. We evaluate our methods and the utility of causal distance on a corpus of research articles taken from the PubMed database.

Portable serverless computing to enable scalable data science

Daniel S. Katz (University of Illinois, Urbana-Champagne) and Kyle Chard (University of Chicago)

Data science depends on data, algorithms/software, computing, and people. In the past, a data scientist would work directly on the computing resource where they wanted to run their algorithms

(e.g., their laptop, a shared cluster, or the cloud). More recently, the data scientist could develop containers with their software and deploy them to the computing resources. Today, serverless computing is an even lighter-weight and easier-to-program option, where only functions are deployed from a development platform to an execution platform. While the architecture that controls the deployment and execution of these functions is often specific to a platform (e.g., Lambda for AWS, Cloud Functions for Google, etc.), we've developed a portable serverless computing platform called funcX (https://funcx.org) that allows a data scientist to portably run their functions on various cloud, local, or HPC resources, including campus clusters and NSF- and DOE-funded HPC systems. These functions can be run where resources are available soonest or cheapest, or where data exists (e.g., a cluster close to an instrument that produces data in real time). By using a cloud-hosted service, functions can be launched in a fire-and-forget style, and can be shared in a user-controllable manner. funcX has been used for problems such as scalable fitting and columnar analysis in High Energy Physics, generating training data for molecular dynamics machine learning, running ensembles of COVID-19 epidemiologic models, real-time experiment analysis at beamlines including Argonne's Advanced Photon Source, scalably executing publicly-stored ML models, and text extraction.

March 9 • Wednesday 8:50am – 9:00am [Announcements](#)
Auditorium

9:00am – 10:00am K [Student Keynote: Using Data Analysis to Solve Human Trafficking](#)
Auditorium

Abstract:
Human trafficking, as defined by the United Nations, is the recruitment, transportation, transfer, harboring or receipt of people through force, fraud, or deception, with the aim of exploiting them for profit. With the increased use of the internet and media, it has been much easier for predators to take advantage of innocent children. Unfortunately, according to the FBI, 57.5% of all juvenile prostitution arrests are Black children. To reduce and fight against trafficking, we used data science techniques to examine ERASE Child Trafficking organization's Facebook page to determine which posts would increase engagement and which were unproductive in having a further reach for their audience. Utilizing the data from the insights provided from engagement, likes, shares, and views, we were able to develop a plan which would increase the outreach and awareness. Using R/RStudio, we analyzed each type of post (picture, text, videos, links, etc.) and the language used to examine engagement via a linear regression model. We will share our data analysis methods and findings that inform how to optimize outreach and overall engagement to solve human trafficking.

Speakers:
Victoria Grase (Clark Atlanta University)
Woomy Michel (Clark Atlanta University)

10:00am – 10:30am [Break, coffee and pastries](#)
Atrium

10:30am – 12:00pm B [Creating a Diverse and Dynamic Learning Environment for the Education of Undergraduate Data Science Students.](#)
Atrium

Speakers: Eric Van Dusen, Brian Wright

Abstract:

University of California, Berkeley and University of Virginia faculty will discuss initiatives being developed or currently underway as it relates to Data Science undergraduate curriculum. Specifically, UC Berkely faculty will provide an overview of how their Data Science Undergraduate Studies (DSUS) was built and leverages open-source tools such as cloud-computing environments on JupyterHub, online textbooks, and automatic grading infrastructure that empowers educators to offer data science and domain-specific data-enabled courses at scale. This will include an overview of the process and decisions that go into building data-enabled curriculums for traditionally non-computational classes. University of Virginia faculty will similarly discuss the development of undergraduate programs through three discrete perspectives: course development, program development, and community-focused research. These perspectives resulted in an emphasis on authentic Data Science course creation, allowance but not dependence on interdisciplinary program development, and the importance of making sustainable relationships with the Charlottesville community to foster hands-on research-driven projects. Throughout the panel, both Universities will include dialogue on how their design approaches work to encourage diversity and inclusion into the rapidly emerging field of Data Science.

Chairs: Brian Wright (University of Virginia) and **Eric Van Dusen** (UC Berkeley)

12:00pm – 1:00pm [Lunch](#)
Dining room/Courtyard

1:00pm – 2:00pm K [Student Keynote: Enabling Exploratory Large Scale Graph Analytics through Arkouda](#)
Auditorium

Abstract:

Exploratory graph analytics is a much sought out approach to help extract useful information from graphs. One of its main challenges arises when the size of the graph expands outside of the memory capacity that a typical computer can handle. Solutions must then be developed to allow data scientists to efficiently handle and analyze large graphs, using machines that have the capacity to handle massive file sizes. Arkouda is a software package under early development created with the intent to

bridge the gap between massive parallel computations and data scientists wishing to perform exploratory data analysis. In this talk, Oliver will highlight the recent contributions made to enable graph analytics in Arkouda, their importance, the challenges that presented themselves, and future steps that need to be taken. This work can help further bridge the gap between high-performance computing (HPC) software and data science to create a framework that is straightforward for all data scientists to use.

Speaker:
Oliver Alvarado Rodriguez (New Jersey Institute of Technology)

2:00pm – 2:30pm Break
Atrium

2:30pm – 4:00pm B Career Development Networking Session
Newport - breakout room

Abstract:
The Career Development Network (CDN) is a group within ADSA that supports the professional development of early to mid-career data scientists. We are very excited to bring together data science practitioners from various fields for a session of introductions, inspiration, learning and connecting. In this session, we will introduce the CDN, its mission, and upcoming plans and activities. We will hear from our previous seed grant recipients about their exciting on-going projects. And, we will brainstorm together on how best to engage our members in the upcoming years and introduce small grant opportunities to help with this effort.

Chairs:
Maryam Vareth (UC Berkeley; UCSF)
Steve Van Tuyl (ADSA)

2:30pm – 4:00pm B Cyberinfrastructure for Data Science
Auditorium

Speakers: David Bader, Daniel Katz
This session focuses on data science uses of cyberinfrastructure for academic researchers. We will discuss the needs of our research community for access to cyberinfrastructure including people, data, software, storage, networks, and high performance computing, and the convergence of big data and high performance computing. Attendees will learn more about national resources and industrial support for large-scale data science, and have an opportunity to share their experiences. As data volumes continue to grow in size, in some cases becoming unmanageable for "laptop science," we will discuss possible pathways for access to cyberinfrastructure and gather community needs. Goals of the

session include capturing any barriers to access cyberinfrastructure resources and documenting "best practices" in the data science academic community.

Chairs: **David Bader** (New Jersey Institute of Technology) and **Daniel Katz** (University of Illinois)

4:30pm – 6:00pm [After party at the Hyatt, Informal discussions on the future of ADSA](#)
The Garden Patio @ the Hyatt Newport Beach

Join us for an after conference cocktail reception and some informal discussions of the future of ADSA. Want to join the Program Committee for the Fall 2022 meeting? Let us know!