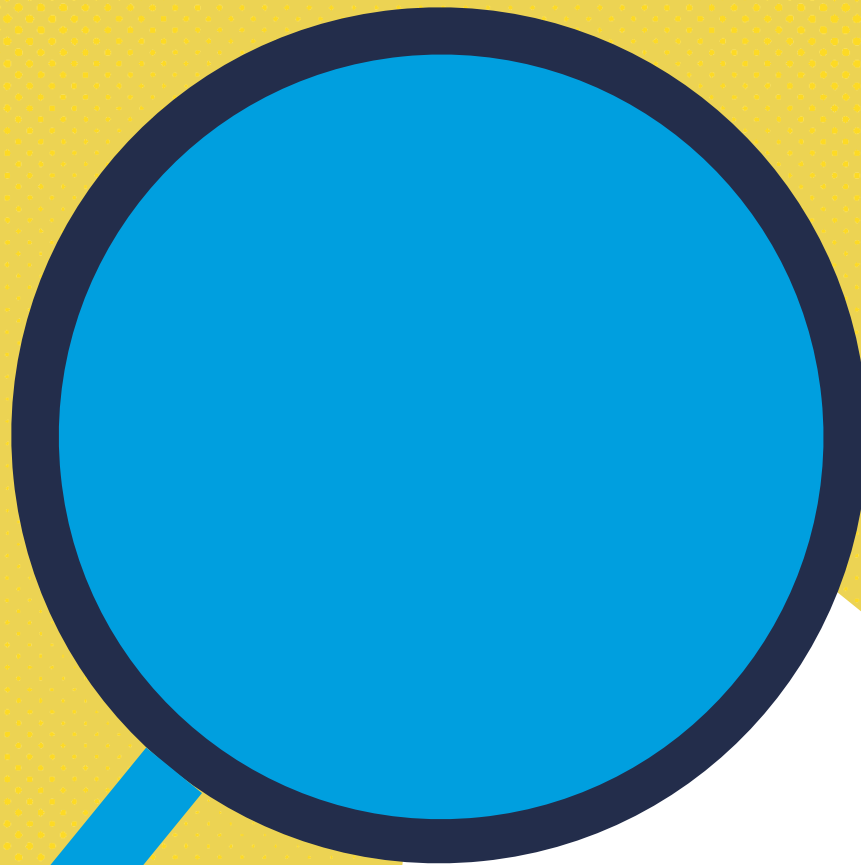


# EXPLORATIONS IN CONNECTION

ANNUAL REPORT

2021  
2022



 UNIVERSITY *of* VIRGINIA

---

**BIOCOMPLEXITY** INSTITUTE

# TABLE OF CONTENTS

<b>ABOUT THE INSTITUTE</b>	<b>3</b>
<b>MESSAGE FROM THE EXECUTIVE DIRECTOR</b>	<b>8</b>
<b>AWARDS AND HONORS</b>	<b>10</b>
<b>THOUGHT LEADERSHIP</b>	<b>12</b>
CONNECTIONS FOR BETTER PREPAREDNESS	<b>14</b>
CONNECTIONS FOR BETTER HEALTH AND RESILIENCE	<b>16</b>
CONNECTIONS FOR SMARTER PROBLEM SOLVING	<b>24</b>
CONNECTIONS FOR STRONGER COMMUNITIES	<b>28</b>
<b>NOVEL TECHNOLOGIES: DIGITAL TWINS</b>	<b>32</b>
<b>EXPERIENTIAL STUDENT LEARNING</b>	<b>38</b>
DATA SCIENCE FOR THE PUBLIC GOOD	<b>40</b>
COMPUTING FOR GLOBAL CHALLENGES	<b>43</b>
<b>ACKNOWLEDGEMENTS</b>	<b>48</b>

# ABOUT THE BIOCOMPLEXITY INSTITUTE & INITIATIVE



# EXPLORATIONS IN CONNECTION

Biocomplexity is about pervasive interaction, adaptation, and co-evolution of the living, technological, and natural worlds. Life, its artifacts, and the natural and built environments make up a complex system of astronomical numbers of possible interactions and relationships. The result is the world we inhabit, the technologies we produce and use, and the co-evolution of these things. Science struggles with the integration of all of this, but the practical reality is that management of ourselves and the systems on this planet is undeniably critical.

The more we learn about the complexities of living and non-living systems, the more we realize how deeply interconnected they are – how transportation systems affect the spread of a virus, how a major company's relocation impacts food security and climate change, how a pandemic impacts gig employment and the economy. Information flows, including deceptive information flows, are a central and challenging aspect of this study. They are not side effects or epiphenomena, rather they are central objects of scientific concern in biocomplexity.

The sheer number of interactions and data makes the study of such scenarios challenging, motivating endless needs for fundamental as well as applied research. But, with a multidisciplinary approach across scientific fields aided by the massive power of computational tools and sophisticated modeling, biocomplexity research helps identify connections, patterns, and approaches never seen before. These insights can be harnessed to expand human perception of our conditions, our intelligence, and our ability to act – and can thus help governments and authorities make better decisions about things that matter to people and to the wellbeing of the planet.

Biocomplexity challenges scientists to rethink their skills, professional focus, work, and workplaces. Biocomplexity can improve leadership at all levels from the working person to the president by providing new tools and concepts to better make sense of the world. Biocomplexity seeks to be part of solutions for the world's most complex and vexing problems. At the Biocomplexity Institute and Initiative at the University of Virginia, this is our mission.

## Our Origin Story

Real-world, large scale problems and their phenomena involve many millions of interconnected and interacting systems that cross or ignore all traditional academic domain boundaries. Science must somehow expand to include integrated issues such as human mobility, climate change, infectious disease spread, social transformation, and the evolution of conflict and warfare, among others. These genuine problems pose deep scientific challenges that require non-traditional teams of scientists with diverse perspectives, expertise, and skill sets coming together in service to society. As much as the scientific research itself, this interwoven, non-disciplinary team approach to complex problem solving has guided the Biocomplexity Institute and Initiative's work for nearly four decades.

### PIONEERING A MULTI-SCIENCE APPROACH

Our story began nearly 40 years ago at the Naval Air Development Center (NADC), a former U.S. Department of Defense research and development center devoted to providing technical superiority to next-generation naval aircraft. In the late 1970s and 80s, advanced sensors, multi-sensor integration, and artificial intelligence-enabled environments with networked control systems became part of the NADC mission.

Moreover, the use of real-time AI systems to support human crews in tactical reasoning had started to gain serious consideration. These aircraft and ship systems were also connected to robotic long-range missiles with adaptive, autonomous capabilities for target search and engagement.

All of this went beyond what was then state-of-the-art in computing, sensing, or AI. In fact, one of the subject aircraft, the F-14, was the first to have a manufactured microprocessor in its design, an innovation driven by the demands and complexity of its operational flight environment. The network-centric control systems were truly human-computing hybrid concepts, not purely biologically (human) or computationally conceived. An integrated human-AI approach was the start of a deep appreciation of the difficulty and beauty of the basic science derived from applied problems and the role of large expert multi-specialty teams in their solution. The first small dedicated technical research team in this area's history came together to study how to integrate a real-time tactical expert system and a sensor data-driven simulator with human aircrew.

A few years later, that work was surprisingly related to problems in the U.S. Strategic Defense Initiative and simulation research projects at the Los Alamos National Laboratory (LANL) in New Mexico. At LANL, multi-science teams have always been part of the mission. A Basic and Applied Simulation Science team formed in the LANL Analysis Division, and several years later, moved to the Computational and Computer Science Division. In this evolution, the group grew significantly to include many graduate students and post-doctoral associates from all over the world.

## ABOUT THE INSTITUTE

In 1991, we developed early concepts and demonstrations of synthetic population representations for simulations of urban vehicular and intermodal transportation, connections of transportation systems to air quality models, and other similar multi-network, multi-scale systems. For 10 years, we led large projects for the U.S. Department of Transportation with the goal of providing groundbreaking capability to understand transportation infrastructure and other decision effects on air quality as required by the Clean Air Act.

Later, in partnership with Motorola and other mobile companies, we developed transportation fows-connected mobile communication networks for the Defense Advanced Research Projects Agency (DARPA). As concerns about terrorism related to critical infrastructure and infectious diseases emerged, we integrated those topics into our work. After the September 11 attacks in 2001, this research was funded under the creation of the National Infrastructure Simulation and Analysis Center (NISAC) for the newly established Department of Homeland Security. Around that time, the National Institutes of Health founded the MIDAS program for infectious disease modeling research and we became a partner in that still ongoing effort. Next, we leveraged our residual research products to begin the Comprehensive National Incident Management System (CNIMS), a comprehensive 15-year research and accelerated development project for the Defense Threat Reduction Agency.

### A TEAM AS DIVERSE AS OUR WORK

A core of Institute research scientists either worked together, began their careers, or both at LANL including: Christopher Barrett, who founded and led this research and the associated groups, Division Directors Sallie Keller, Christian Reidys, and Madhav Marathe, Deputy Director Stephen Eubank, and Professors Achla Marathe, Anil Vullikanti and Henning Mortveit. This intense collaborative work environment and commitment to a multi-disciplinary science program served as the catalyst for decades of state-of-the-art changing research.

What began at LANL eventually joined the Virginia Bioinformatics Institute at Virginia Tech in 2004 and 2005. With access to the various disciplines, the research itself, funding, and a perpetual pool of research students, the university setting proved ideal for the study of biocomplexity. One of the most important additions to biocomplexity that was added at Virginia Tech and remains, is core capability in bioinformatics. Today, we operate two national bioinformatics data and bioinformatics analytics centers, a NIH-funded center for viral and bacterial genomes now operating for nearly 20 years, and a new Centers for Disease Control and Prevention (CDC) center for pathogen genomes.

## ABOUT THE INSTITUTE

Our team has witnessed and served practice at the national level during some of the world's most consequential events over the last four decades: the transformation of state-sponsored conflict, the end of the Cold War, the beginning of the World Wide Web, the beginning and end of one of the largest national security programs since the Manhattan Project, two major land wars, numerous smaller military engagements, September 11 and the establishment of Department of Homeland Security, and numerous major infectious disease outbreaks including the COVID-19 pandemic. Like the living systems we have studied, our team itself has become a complex system of interacting parts. Social and behavioral scientists, mathematicians and biologists, statisticians and computer scientists, and others form the Biocomplexity Institute.

### NEW BEGINNINGS AND A BRIGHT FUTURE

In 2018, we established our new home at the Biocomplexity Institute and Initiative at the University of Virginia (UVA). In four short years, the Institute has grown to nearly 135 faculty, students, and staff. Since joining UVA, we have authored over 500 published works, worked with major institutions including the U.S. Census Bureau, the U.S. Departments of Defense and Agriculture, and the CDC, and received more than 135 funding awards totaling over \$60.25 million from partners including the Bill and Melinda Gates Foundation, National Science Foundation (NSF), and National Institutes of Health (NIH).

For much of the last four decades, we have worked to anticipate, recognize, and respond to some of society's most urgent and complex problems, and have made major contributions both domestically and internationally. As we look to the future, we will continue exploring connection and complexity and diving into the unknown. The world's problems will continue to guide our direction.



MESSAGE FROM THE EXECUTIVE DIRECTOR

# CHRIS BARRETT



Dear Biocomplexity Institute,

We are completing our fourth year together here at UVA! This is the first annual report of its kind that we have produced at the University of Virginia and it documents 2021-22 as something of the Biocomplexity Institute and Initiative’s coming of age. A new beginning. Make no mistake though, this is just the beginning. The future you are now creating will be more remarkable than this excellent beginning place you have made so far. We are just starting.

Biocomplexity is a general non-disciplinary organizing scientific and engineering approach to comprehension, design, and management of the layers of detailed interaction related to the subtle tapestry of living, technologically and socially extended systems. We are all both creating and created of this tapestry. Biocomplexity problems motivate fundamental research questions that are deep and beautiful while also serving the immediate practical, often messy, present. Individual physical, biological, behavioral, and social sciences hold no particular claim to any of the composite problems that emerge. Conceiving, making, and employing analytical tools and methods to help make sense of all that, a few steps at a time with our more disciplinary colleagues in valued collaborations, is the scholarly business of the Institute. Committing to providing that work to meet the real needs of others is our service responsibility.

Our excellent research teams and equally unique and excellent research support teams have contributed tirelessly, practically, and as engaged scholars, to the actual week-to-week management of difficult current events at the community, regional, and national levels. You have, while creating this Institute and getting it off the starting line, simultaneously served as something of an active analytical research hub and national complex systems analysis resource. While in academic and practical service, you have produced several hundred nationally and internationally recognized scientific and scholarly peer-reviewed publications, provided hundreds of regular analytical reports and briefings to national and regional leaders, given professional service, written and successfully submitted over 200 proposals, produced young scholars in the tradition of the Institute, and created a unique administrative and operational organization in every detail – from nothing, all at once.

There is an old blessing, wishing one may not live in interesting times. But that, obviously, has not been your fate so far during these first years here. Without exaggeration, it has been a historic period since our arrival; one that has demanded a great deal from each of you, our teams, and our institution. And yet, you have found the capacity to create something unique and exceptional at the University of Virginia. You have created a tradition and legacy of excellence in the conduct of thematically driven, problem-solving fundamental research and set a direction for sustaining and growing this work tradition going forward.

In recognition of the achievements we have accomplished together, this annual report is dedicated to you.

Chris Barrett  
Executive Director and Distinguished Professor in Biocomplexity, Biocomplexity Institute  
Professor of Computer Science, School of Engineering and Applied Science

## Abbe Grant Recipient, Heidelberg Laureate Forum



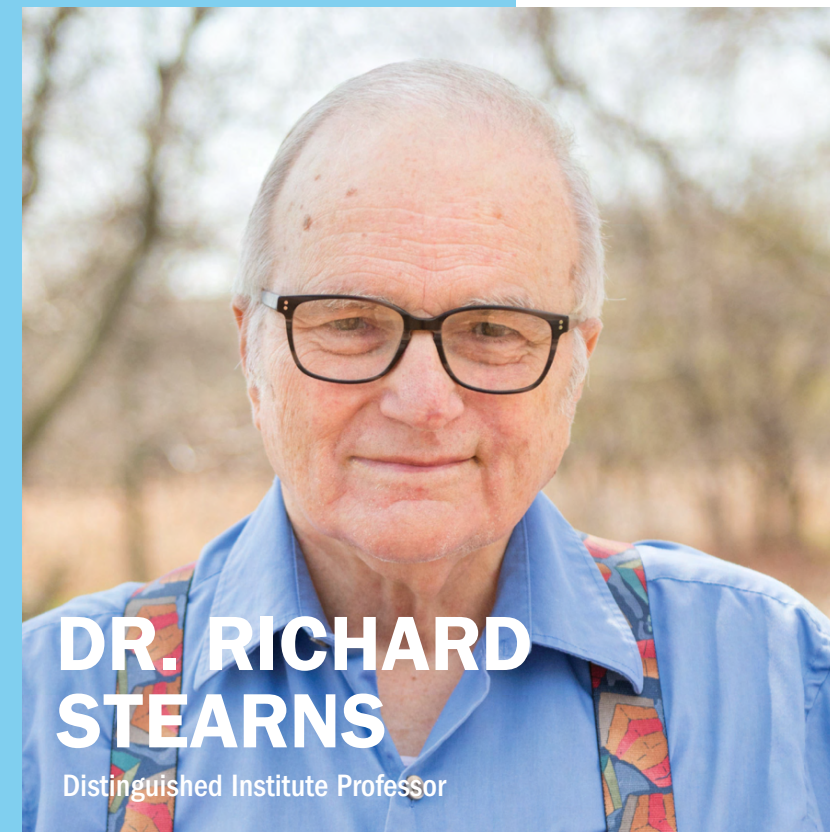
Dr. Baltazar Espinoza Cortes was one of only 30 researchers worldwide to receive the Abbe Grant to attend the 9th Annual Heidelberg Laureate Forum (HLF) in September 2022. A Research Assistant Professor with a background in Mathematical Epidemiology, Dr. Cortes was selected for the grant through a highly competitive process based on diverse academic and achievement criteria.

The Heidelberg Laureate Forum (HLF) is an annual networking conference where 200 carefully selected young mathematics and computer science researchers convene with laureates of the two disciplines, including Nobel Prize, Abel Prize, ACM Alan M. Turing Award, ACM Prize in Computing, and Fields Medal winners. The event provides young researchers the opportunity to gain feedback on their areas of research from the fields' most notable and distinguished experts.

As a recipient of the ACM Alan M. Turing Award (1993), Distinguished Institute Professor Richard E. Stearns has an open invitation to attend the HLF annually and the privilege of nominating young researchers as potential attendees. Professor Stearns nominated Dr. Cortes and Dr. Lijing Wang, who was unable to attend due to scheduling conflicts.

Distinguished Institute Professor Dr. Richard E. Stearns received the prestigious ACM Alan M. Turing Award in 1993 in recognition of his groundbreaking joint research with Professor Juris Hartmanis of Cornell University, which established the field of computational complexity. He has made pioneering contributions to several different areas of computer science including theory of computation, formal languages, compilers, analysis of algorithms, database systems, and game theory.

Dr. Stearns also holds the title of Distinguished Professor Emeritus in the Department of Computer Science at the University of Albany – State University of New York (SUNY), where he spent 22 years, including seven years as the Department Chair. He has held additional academic appointments at several institutions including Hebrew University in Jerusalem; Mathematical Sciences Research Institute in Berkeley, California; and Rensselaer Polytechnic Institute in Troy, New York. He began his professional career in 1961 and spent 17 years with General Electric Research Laboratory, now known as GE Global Research, based in Schenectady, New York.



## ACM Alan M. Turing Award Recipient 1993



# THOUGHT LEADERSHIP

# CONNECTIONS FOR BETTER PREPAREDNESS

## A Breakthrough for Faster, More Accurate Threat Detection

### *Mathematical Biocomplexity Develops Novel Molecular Evolution Framework*

Breakthroughs are only possible when we challenge the status quo.

In the ongoing fight against rapidly mutating viruses such as COVID-19 and monkeypox, genomic surveillance has become a critical tool. Traditionally, scientists evaluate massive amounts of genomic data to group similarities and flag distinctions in the viral multiple sequence array (MSA) that may signal the emergence of new mutations that could pose an imminent public threat. Once detected, scientists and public health officials use these findings to facilitate early awareness and develop effective mitigation methods for emerging variants.

In 2022, Mathematical Biocomplexity researchers challenged the status quo with a novel approach. Rather than identifying mutations across viral MSA sequencing data, the team looked at the MSAs as individual structures of sites, and identified synchronized mutational patterns among them. The idea is that if a group of sites is inherently linked or related, then they will mutate simultaneously becoming a part of what is called the MSA's motif complex. Therefore, researchers look for maximal sets of synchronized sites, which signal an abundance of correlated mutational activity and molecular diversity.

**“Using the motif complex concept, we have developed a surveillance tool that can detect viral variants of concern much faster and issue alerts much earlier than current methods,”** said Mathematical Biocomplexity Division Director Christian Reidys.

“Within this framework, we don’t need any biological knowledge of the virus – we focus solely on mutation synchronization signatures. When the configuration of change is significant, we know that something big is happening.”

To test the motif complex methodology, the Mathematical Biocomplexity team conducted a retrospective analysis of COVID-19 sequence data collected from November 2020 through August 2021. Applying their motif-centered surveillance tool to this data, they detected emerging variants of concern notably faster and with an accuracy of over 85 percent, enabling alert generation within just hours of acquiring the relevant data.

“The framework’s benefits are numerous – certainly speed, accuracy, and lack of cost. Further, it is unbiased. It is looking only at synchronization patterns – the ‘patterns that survive’ or the ‘winners.’ It doesn’t really care about biology, which means its application is very broad. We’ve utilized it on COVID-19 and monkeypox data, but the method is applicable across a multitude of disciplines and data sources, really wherever synchronization matters.”

Reidys explained that the motif complex concept originated from pure math work on weighted complexes. This breakthrough, with its transdisciplinary application and relevancy, is a quintessential example of the science of biocomplexity and its benefit to humankind.

## LEARN MORE

[Buying Time: Detecting Vocs in SARS-CoV-2 via Co-evolutionary Signals](#)

[Rapid Threat Detection in SARS-CoV-2](#)

[The Arithmetic Topology of Genetic Alignments](#)







# CONNECTIONS FOR BETTER HEALTH AND RESILIENCE

THOUGHT LEADERSHIP

## From Hurricanes to War: Data is Vital to Building More Adaptable, Resilient Communities

How does saltwater intrusion from climate change impact our nation's coastal farmland and housing? How does the threat of crime factor into the decision to evacuate before a hurricane? How has the Russian invasion of Ukraine impacted access to healthcare for displaced Ukrainians with chronic illness?

These questions along with many of our world's most complex problems involve interdependent networks and infrastructures. The Biocomplexity Institute is one of the nation's leaders in untangling and streamlining these networks to extract data enabling evidence-based, science-backed decision making in the face of disasters and emergencies.

The Institute's Network Systems Science and Advanced Computing (NSSAC) division applies high-resolution scalable models and an interaction-based simulation approach to study connections and interdependencies among societal infrastructures such as transportation systems, power grids, and communication systems and the impact of various physical, technological, informational, and human factors. Below is a sampling of the NSSAC division's work related to resilience, disasters, and sustainability over the last year:

### **Coastal Futures: Building Capacity for Data-Driven Adaptation in Rural Coastal Communities**

In partnership with UVA's Environmental Resilience Institute (ERI) and the Biocomplexity Institute's Social and Decision Analytics division, the NSSAC is in its second of a five-year, National Science Foundation-funded project to inform stakeholders in coastal Virginia communities on the impacts of various environmental effects of climate change such as saltwater intrusion and flooding on agriculture, drinking water, habitat loss, infrastructure, and more. The NSSAC is developing an open-source data, modeling, and visualization platform to enable stakeholder collaboration and equitable decision making to create more adaptive communities.

## THOUGHT LEADERSHIP

### From Space to Front Porch: Connecting Earth Observations to Health Outcomes with an Environmental Exposure Modeling System

In collaboration with Virginia Tech and Johns Hopkins University and funded by NASA, the Institute is studying how extreme weather along with geography, the built environment, and human activity patterns affect human vulnerability. The NSSAC team is developing models based on Earth Observation data, population models, health data, and the CDC's Social Vulnerability Index (SVI) along with synthetic populations and disaster simulations to assess myriad disaster scenarios. For example, by simulating hurricane evacuation patterns, the team can estimate which people are most vulnerable to hazards such as flooding, drowning, water contamination, disease transmission, depression, and PTSD, among others.

### Quantifying Distributional Health Damages of Extreme Weather Events

In August 2022, the Institute embarked on a new NASA-funded project to improve estimation of health damages due to flooding, heat, and cold waves. Integrating Earth Observation data and synthetic population modeling of human movements, the NSSAC team aims to improve estimation of human proximity and exposure to extreme temperatures, flooding, and air pollution during and following extreme weather events. While this work has numerous applications in disaster management and public health, it will immediately serve to better inform health damage estimates and accounting methods for updating the social cost of carbon (SCC), a metric designed to quantify and monetize climate damages, a priority for the Biden-Harris Administration.



**Our advanced population modeling and simulation capabilities allow us to connect climate data with health data to quantify the impacts of climate change and disasters. These projects are leading to a more detailed understanding of climate-induced social vulnerabilities and pathways to their mitigation.**

**– Samarth Swarup**  
Research Associate Professor,  
Network Systems Science and  
Advanced Computing



## THOUGHT LEADERSHIP

### Critical Resilient Interdependent Infrastructure Systems and Processes 2.0 (CRISP 2.0)

The CRISP 2.0 project funded by the NSF is focused on better understanding natural disasters and improving human resilience. The Institute's NSSAC team is overseeing three major components. First, to ultimately increase evacuation rates in an impacted region, the team has developed models to study hurricane evacuation decisions and the impact of peer effects and fear of property loss on those decisions. Second, the team is developing a web application that enables citizens to share information about goods and services to improve their abilities to adapt amidst a natural disaster. Finally, NSSAC is studying techniques for citizens to share resources in the wake of an emergency.

### Disease Burden Among Ukrainians Forcibly Displaced by the 2022 Russian Invasion

In partnership with National Institutes of Health, the Yale School of Public Health, and the Public Health Center of the Ministry of Health of Ukraine, among others, the NSSAC division is part of the first study to estimate the burden of some of the most prevalent chronic diseases such as cancer, diabetes, and tuberculosis among forcibly displaced Ukrainians due to the 2022 Russian invasion. Combining a spatiotemporal model of forcible displacement with age- and gender-specific estimates of these various diseases in each of Ukraine's 629 raions (districts), the NSSAC team is estimating disease burden in Ukrainians displaced both within Ukraine and to other countries. The findings will identify regions in which healthcare systems may face immediate strain and suggest critical targets for humanitarian aid within Ukraine.

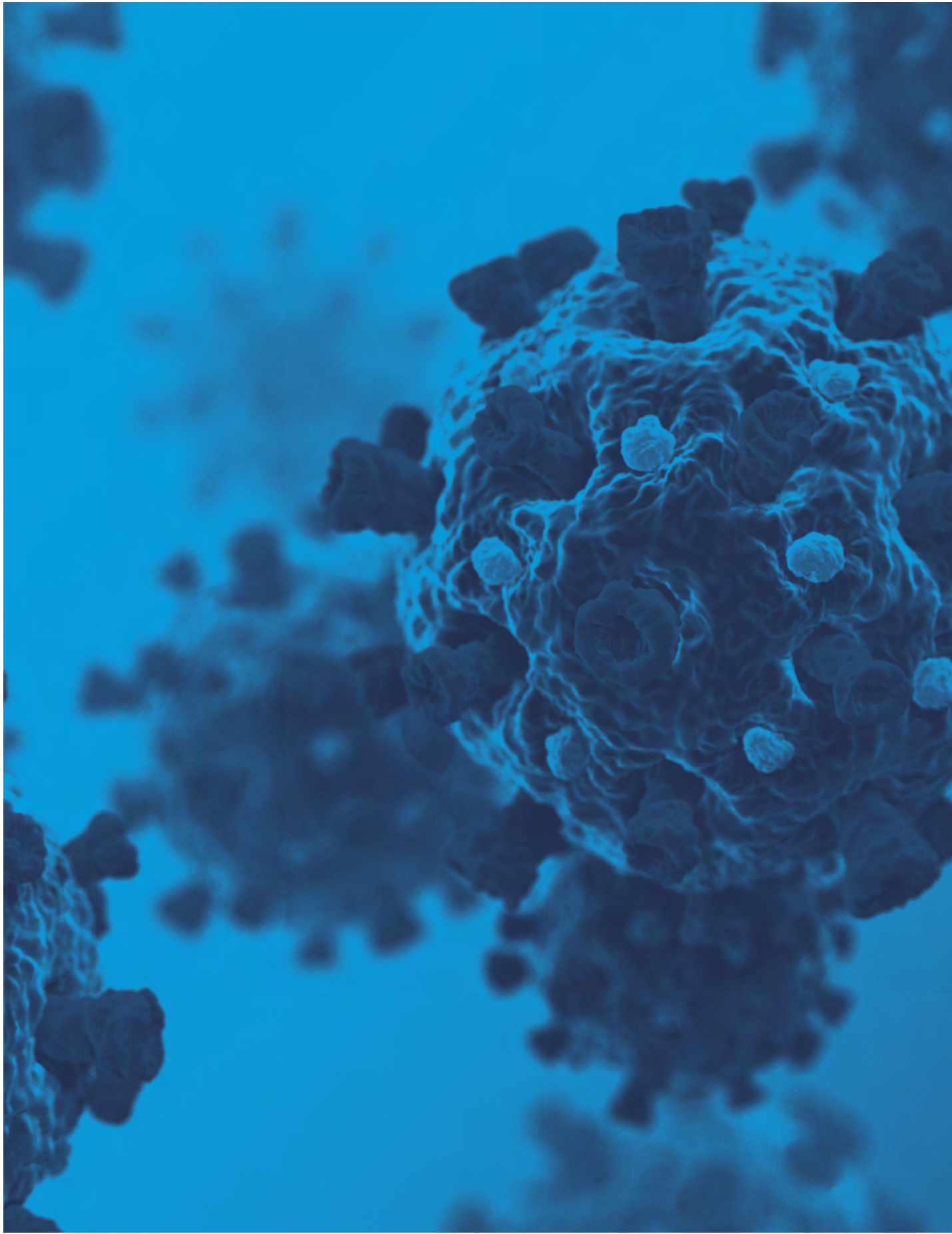
**“Our teammates at Florida International University produced survey data indicating that families who wish to evacuate in the face of oncoming natural disasters may not due to concerns about crime and theft. We developed a simple model of this behavior to estimate evacuation levels in large cities such as Houston and Miami, contrasting those results with results from classic models of evacuation. We demonstrated that evacuation rates may be as much as 50% less when accounting for the influence of crime. We are currently using these results to study resiliency of those families that shelter in place during a natural disaster.”**

– Christopher Kuhlman, Research Associate Professor, Network Systems Science and Advanced Computing

## LEARN MORE

[NSSAC Resilient Societies and Interdependent Infrastructures Area of Focus](#)

[Natural Disaster Evacuation Modeling: The Dichotomy of Fear of Crime and Social Influence](#)



## THOUGHT LEADERSHIP

# A Global Leader in Pandemic Planning and Response

### Institute Innovation Is Transforming How We Prevent, Control, and Respond to Disease Threats

The Biocomplexity Institute has been at the forefront of epidemiological study and modeling since 2001, and has become a critical partner to local, state, and federal government officials in their response to the COVID-19 pandemic. The Institute research team is a global leader in producing innovative computational models and real-time simulations using high-performance computing (HPC) for epidemiological forecasting. Using these simulations, the Institute informs data-driven, science-based decision making for partners including the Commonwealth of Virginia, Centers for Disease Control and Prevention (CDC), National Science Foundation, and U.S. Department of Defense.

Over the last year, the Institute's Network Systems Science and Advanced Computing (NSSAC) division has supported pandemic planning and response in three broad areas: forecasting and supercomputing, mobile vaccine placement, and biosurveillance.

#### Forecasting and Supercomputing

Since the earliest stages of the pandemic, Institute researchers have employed HPC and computer simulations to forecast its evolution. Forecasting has served as an essential early warning tool for communities across the United States, enabling better preparedness and response for everyone from citizens and policy makers to healthcare workers.

The availability of near real-time pandemic incidence data has been a tremendous asset in understanding its dynamics. Using this data, the Institute team developed a framework to generate granular level-projections to more accurately forecast when, where and how severe outbreaks would likely be.

**“The ongoing support provided to the commonwealth and nation has been an excellent application of the tools we’ve developed over the years, and we are heartened to see them have real-world impact.”**

**–Bryan Lewis,  
Research Associate  
Professor, Network Systems  
Science and Advanced  
Computing**

## THOUGHT LEADERSHIP

“Using our live individual-scale modeling, we can simulate where COVID is going in every county of the U.S. every week,” NSSAC Division Director Dr. Madhav Marathe said. “From our knowledge, we are the only team who does agent-based modeling that directly supports decision making in Virginia and elsewhere in the country. These models provide a virtual laboratory to carry out counter-factual experiments – such experiments cannot be done in the real world.”

Using multi-scale models, the Institute team contributed to the CDC Scenario Modeling Hub (SMH) and forecasting hubs. The SMH results have been routinely briefed to national COVID response teams and senior leadership at the CDC, and presented at several rounds of the CDC’s Advisory Committee on Immunization Practices (ACIP) meetings which ultimately approved and recommended COVID vaccinations.

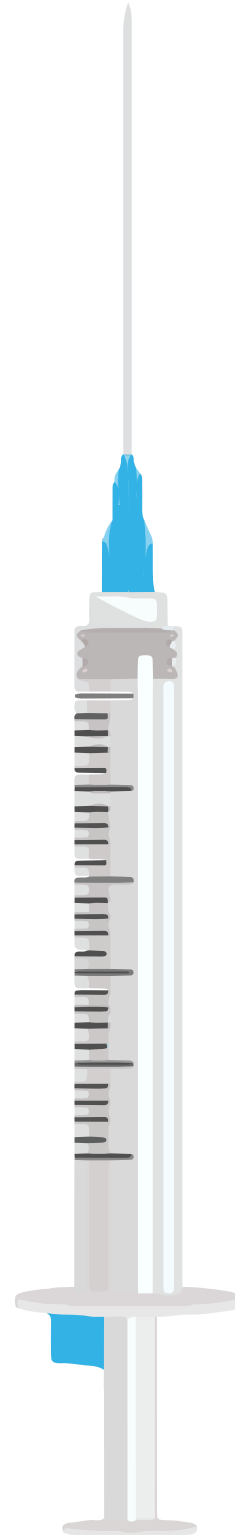
The Institute team garnered national acclaim for this work in late 2021 being named one of five finalists for the ACM Gordon Bell Special Prize for High Performance Computing-Based COVID-19 Research. While the prestigious ACM Gordon Bell Prize has been awarded annually since 1987, a special prize was established in 2020 for outstanding research using HPC in innovative ways to deepen understanding of the nature, spread, or treatment of COVID-19.

### Mobile Vaccine Deployment

Over the last year, the Institute has worked closely with the Virginia Department of Health (VDH) to develop a dynamic, data-driven model for mobile vaccination site placement targeting under-vaccinated populations. When the VDH initially started the program, officials made educated guesses when deploying mobile vaccination sites to communities with diverse populations and high rates of infection. After a strong initial response, attendance dropped drastically. Knowing that data could help better inform placement of the mobile units, officials reached out to the Institute for help.

In collaboration with VDH and a team led by Jure Leskovec at Stanford University, the NSSAC team developed sophisticated models integrating anonymized cell phone data showing where and when Virginians were traveling and the impact of pandemic safety restrictions. Next, they integrated specific census and vaccination uptake data to identify locations most likely to be visited by unvaccinated and vaccine-hesitant people. The idea was to expand vaccination rates by meeting people where it is most convenient for them.

Based on computer-generated models, the Institute provided weekly updates to the VDH with the top 25 recommended locations in designated health districts for mobile clinics. Utilizing this data, the commonwealth improved access to vaccines and vaccination rates for Black and Latino residents over the course of a few months.



## THOUGHT LEADERSHIP

### Biosurveillance

Since 2020, the Institute has worked in close partnership with the Commonwealth of Virginia and VDH to establish and expand biosurveillance to efficiently identify new COVID variants, how the disease is transforming, and where it might be going next. This work laid the foundation for the Institute’s inclusion in a newly announced five-year project funded by the CDC to establish the Pathogen Genomics Centers of Excellence (PGCoE) network.

The network will improve innovation and technical capacity in pathogen genomics, molecular epidemiology, and bioinformatics to better prevent, control, and respond to microbial threats of public health importance. The network will include five centers across the country, each consisting of a health department and one or more academic institutions.

The Virginia Division of Consolidated Laboratory Services will include the UVA Biocomplexity Institute and Initiative, UVA School of Medicine led by Dr. Amy Mathers, VDH, and Virginia Commonwealth University. The network will also include centers in Georgia, Massachusetts, Minnesota, and Washington.

“The Centers of Excellence will enhance technology and expertise in public health so we can have the best understanding of existing and future disease burden,” Marathe said. “The Institute has been working to create a clear picture of the COVID-19 pandemic so decision makers can allocate resources and respond more effectively, and this represents a continuation of that effort.”

**“UVA has been working to create tools to better interpret biosurveillance to promote the health and security of Virginians. Our collaboration with the Virginia Department of Health aims to create algorithms for deployment design to better estimate disease burden, minimize uncertainty for disease forecasting, and further characterize the hallmarks of pathogens of concern.”**

- Andrew Warren, Research Assistant Professor, Network Systems Science and Advanced Computing

## LEARN MORE

[UVA Biocomplexity Institute COVID-19 Pandemic Response Website](#)

[“Anonymized Cell Phone Data is Helping Virginia’s Local Health Officials Select Mobile Vaccine Sites”](#)

[Data-Driven Real-Time Placement of Mobile Vaccination Distribution Sites](#)

[CDC Announces \\$90M Funding to Support New Pathogen Genomics Centers of Excellence Network](#)

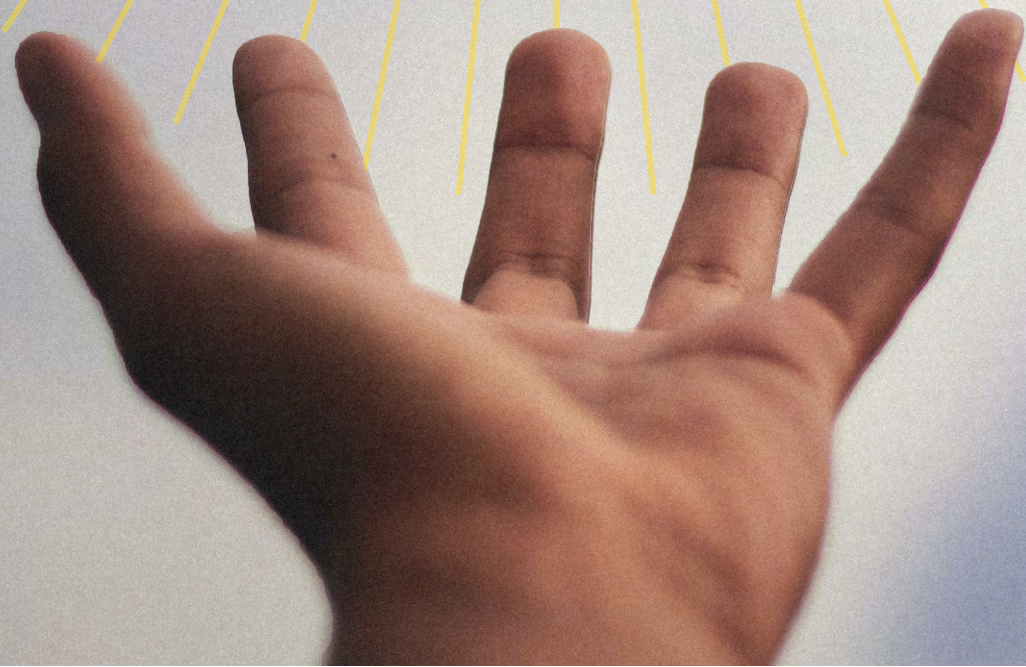
[“Fighting COVID-19 with HPC”](#)

CONNECTIONS FOR

SMARTER



PROBLEM SOLVING



## THOUGHT LEADERSHIP

# A BOLD NEW VISION FOR THE CENSUS BUREAU

### Biocomplexity Institute Proposes Shift in Focus to Data Purpose and Use

The United States census has played a significant role in our democracy since its founding. When the Founding Fathers penned the U.S. Constitution in 1776, they included a mandate for a census every 10 years to accurately apportion the House of Representatives, which laid the foundation for a federal statistical system.

The first census was carried out in 1790 when federal marshals visited every household in the newly formed United States on horseback. Since then, the census has evolved into a highly sophisticated, multi-million dollar endeavor to keep pace with the country's extraordinary growth and advancement. Today, federal agencies rely on census-derived data for House apportionment and to allocate more than \$1.5 trillion annually to local, state, and tribal governments for public health, education, transportation, and other essential needs.

Since officially opening its doors in 1902, U.S. Census Bureau's mission expanded well beyond conducting the decennial census. Now, it serves as the nation's leading provider of comprehensive, quality data about its people, places, and economy. To achieve this mission, the Bureau has long maintained a commitment to innovation in both technology and data collection, which has driven the advancements from manual coding and tallying of self-enumeration and census-by-mail to now computer-based and Internet-enabled technology, among many others.

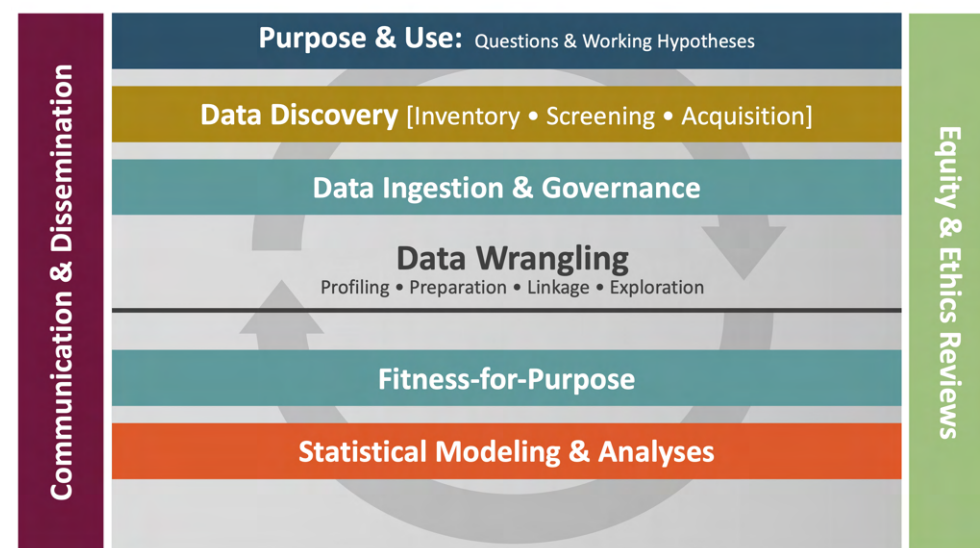
For its next iteration, the Census Bureau is ready to move forward in the 21st century to take advantage of data sources and scientific innovations. In 2020, the Bureau looked to the Biocomplexity Institute and Initiative to develop a vision for a "21st Century Census Curated Data Enterprise." The Curated Data Enterprise (CDE) is both an infrastructure and a continuously evolving ambition to shift the focus from individual data elements and surveys to the purpose and use of the data instead. The vision is to collect, integrate, and maintain data from a variety of public and private sources to empower the Bureau and its data users to develop smarter, timelier, and more comprehensive measures of America's demographic changes, social trends, and economic activity.

## THOUGHT LEADERSHIP

Social and Decision Analytics Division Director and Distinguished Professor in Biocomplexity Sallie Keller said, “Over the last few years, the Institute has done extensive work linking data from diversified sources to analyze meaningful information and trends at the local level. Building on this concept, we are motivating a paradigm shift for the Bureau – to think about how we can connect census-derived information with outside government and third-party data for broader purposes and uses. The CDE will enable us to expand data integration and analysis to study complicated, pressing issues like emergency preparedness, broadband deployment, and climate change.”

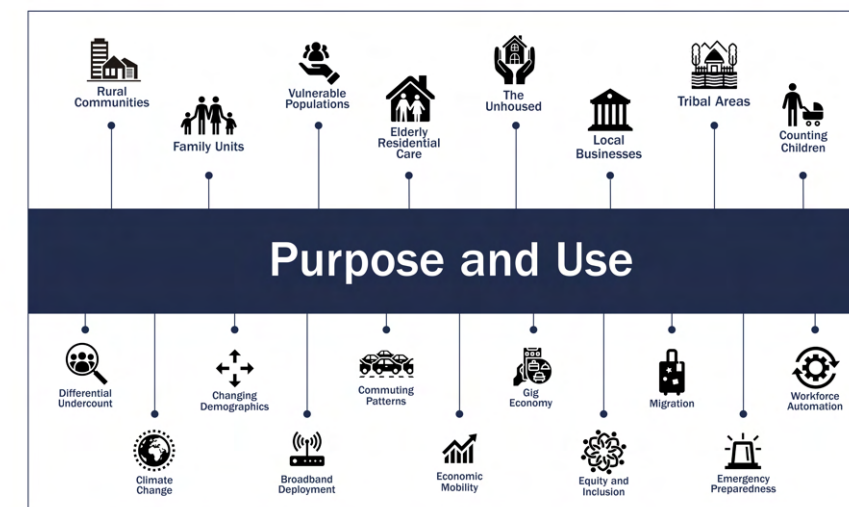
The Institute developed a rigorous, transparent, repeatable CDE framework. Given a purpose, this framework creates a scaffold to move through a series of steps starting with problem identification through statistical modeling and analysis. The framework requires continuous communication and dissemination of results to stakeholders, as well as equity and ethics reviews to ensure integrity throughout the process.

With support from both the Sloan Foundation and the Census Bureau, the Biocomplexity Institute shared the CDE concepts with a diverse set of the Bureau’s key stakeholders, including researchers, economic developers, public policy analysts, economists and business leaders, and others. This year, the Institute published “A 21st Century Curated Data Enterprise: Spring 2022 Report,” which presents their expert viewpoints on the effort, including three comprehensive application scenarios. Collectively, the community concluded that the CDE is a timely evolution beyond the survey-only model that has reached its scientific and practical limits in an era of declining response rates and an overwhelming need for more data, more often, more quickly.



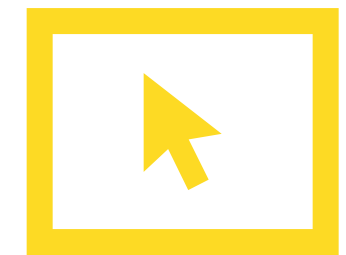
## THOUGHT LEADERSHIP

“As we advance into the 21st century, we are **experiencing increased demand for our data**, struggling with challenges to traditional data collection methods, and exploring rich new data sources and tools that can **revolutionize what we do and how we do it**,” said Ron Jarmin, Census Deputy Director and 2021-2022 Acting Director. “Our success critically depends on our **ability to seize the opportunities** in front of us to deliver statistical products that **address the increasingly complex and diverse needs of our users**.”



Essential to this modernization are **reforms designed to integrate or connect data** across administrative records, surveys, and other public and private data sources into a **Curated Data Enterprise (CDE)**. The linking of now disconnected survey frames on businesses, housing units, jobs, and people into a unified resource forms the foundation of the CDE concept. To that end, the CDE is best described as both an **infrastructure and a continuously evolving ambition** to empower and enable Census Bureau scientists and their data users to **develop together new measures of the nation’s people, places, and economy**.

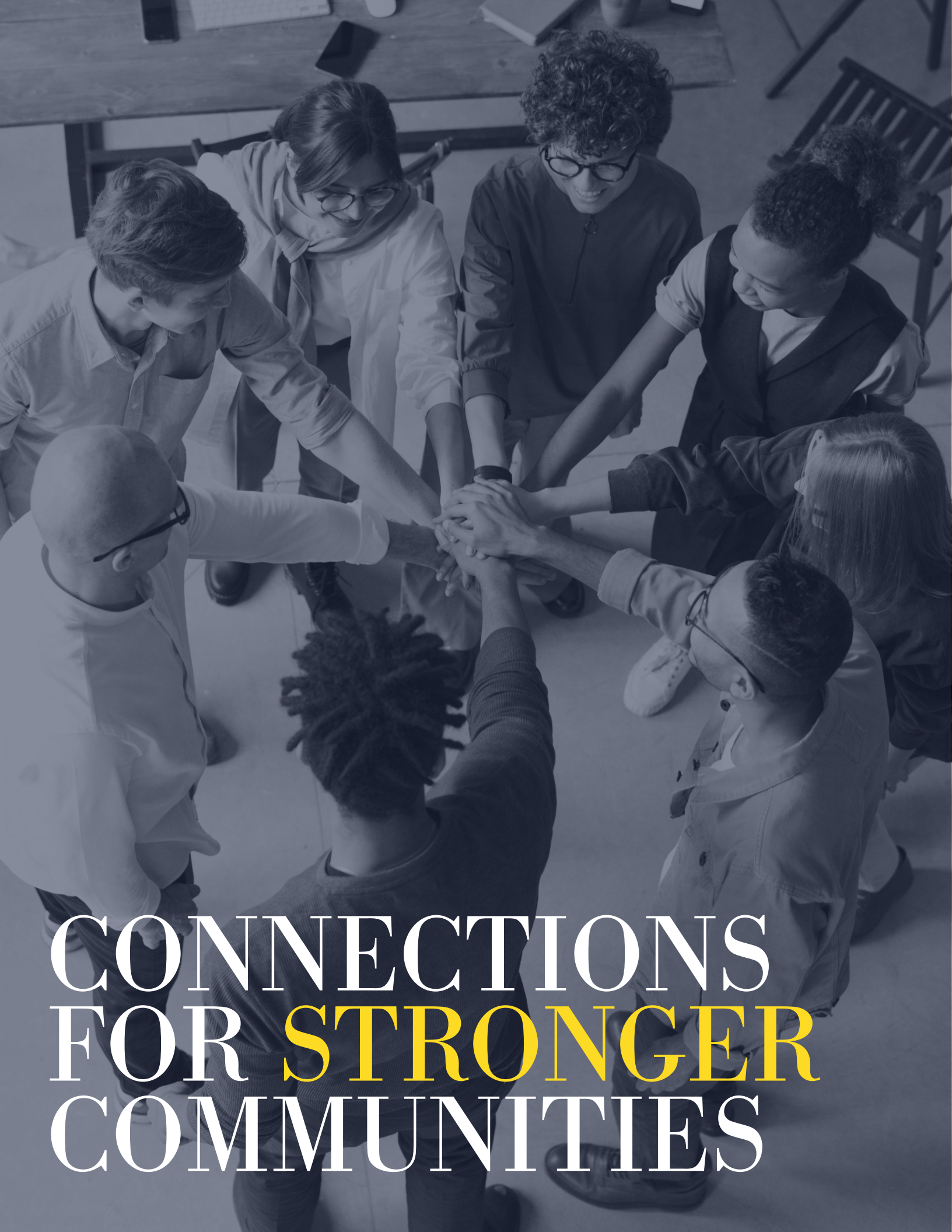
## LEARN MORE



A 21ST CENTURY  
CENSUS CURATED  
DATA ENTERPRISE



A 21ST CENTURY CURATED DATA  
ENTERPRISE REPORT: SPRING 2022



# CONNECTIONS FOR STRONGER COMMUNITIES

## THOUGHT LEADERSHIP

### Putting Data to Work for Smarter, More Equitable Growth

#### *Social Impact Data Commons Supports Equity-Informed Decision Making Across the Northern Capital Region*

To make evidence-informed decisions, communities across the United States require a wide variety of data ranging from demographics and social behaviors to resource availability, health trends, and much more. This data must be timely and topical to the decisions at hand and rendered in geographies relevant to local decision making. However, these communities often lack the sophisticated analytical tools necessary to collect, transform, and curate these data. The Biocomplexity Institute and Initiative is on a mission to change that.

The Institute's Social and Decision Analytics division is leading a cross-disciplinary team to develop an open curated knowledge repository that co-locates data from a variety of sources, including public data-sets, administrative records, and private entities. The Social Impact Data Commons uniquely integrates much-needed analytical tools designed to track issues over time and geography, enabling governments and community stakeholders to continuously learn from their own data and make decisions based on real-time, applicable information. The division is continuing its partnership with the Collaboratory for Applied Data Science in Business (jointly supported by the Darden School of Business and School of Data Science) and the Engineering Systems and Environment Department for the project.

In today's technologically advanced world, we have access to an inordinate amount of complex data. But, it is not the amount of data that is important. Instead, it is how we use it that matters.

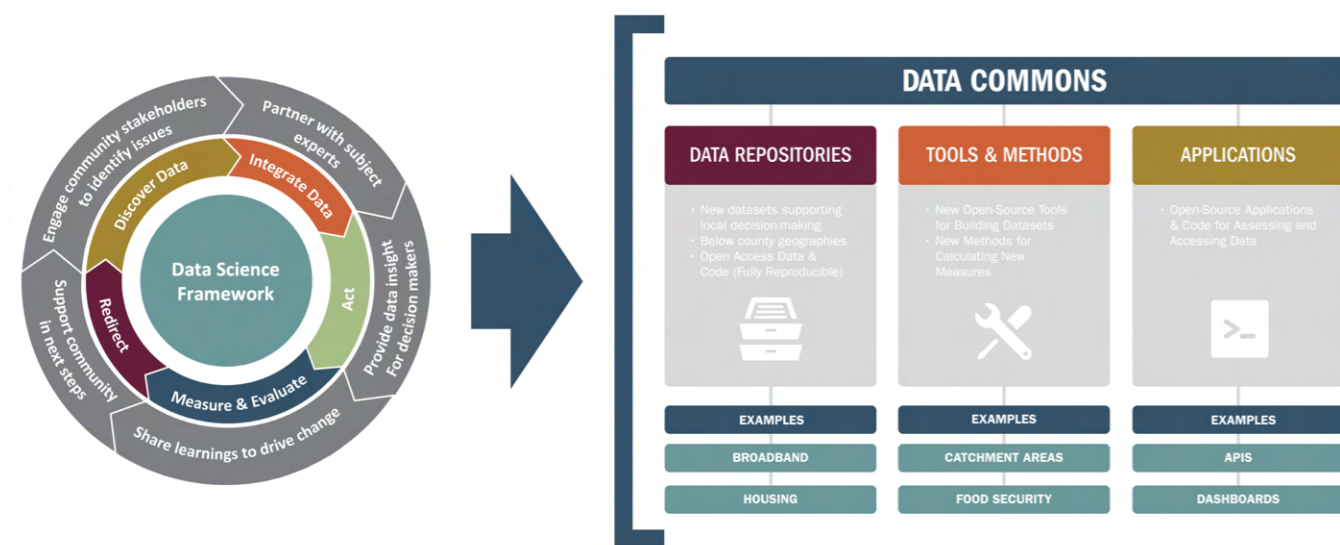


## THOUGHT LEADERSHIP

Development commenced in fall 2021 with a focus on providing these data and tools to each locality in the Washington, D.C., metropolitan region, collectively referred to as the Northern Capital Region or NCR. The goal is to build a data-driven resource to inform equitable and sustainable growth – a pressing issue for a region that is already home to more than 5.5 million people with sustained growth expected over the next 25 years.

The Data Commons will be especially useful with major local events such as the arrival of Amazon’s new headquarters (HQ2) in Northern Virginia, which is expected to bring an influx of 25,000 new and thousands more indirect jobs in the next decade. While the new headquarters will drive welcome economic growth, it also has the potential to exacerbate many long-term concerns for the region, such as food insecurity, climate change, and lack of affordable housing. The Data Commons connects a variety of users to data tracking these local effects along with the analytical tools to inform decisions with widespread public impact.

In the first phase of work, the team, led by Aaron Schroeder, principal investigator and research associate professor in the Social and Decision Analytics Division, established nine data repositories within the Data Commons, an operational dashboard, APIs, and a set of tools to support these entirely new datasets. Within its initial year of development, the Data Commons has already made an impact toward its intended goal.



SOCIAL IMPACT DATA SCIENCE FRAMEWORK

## THOUGHT LEADERSHIP

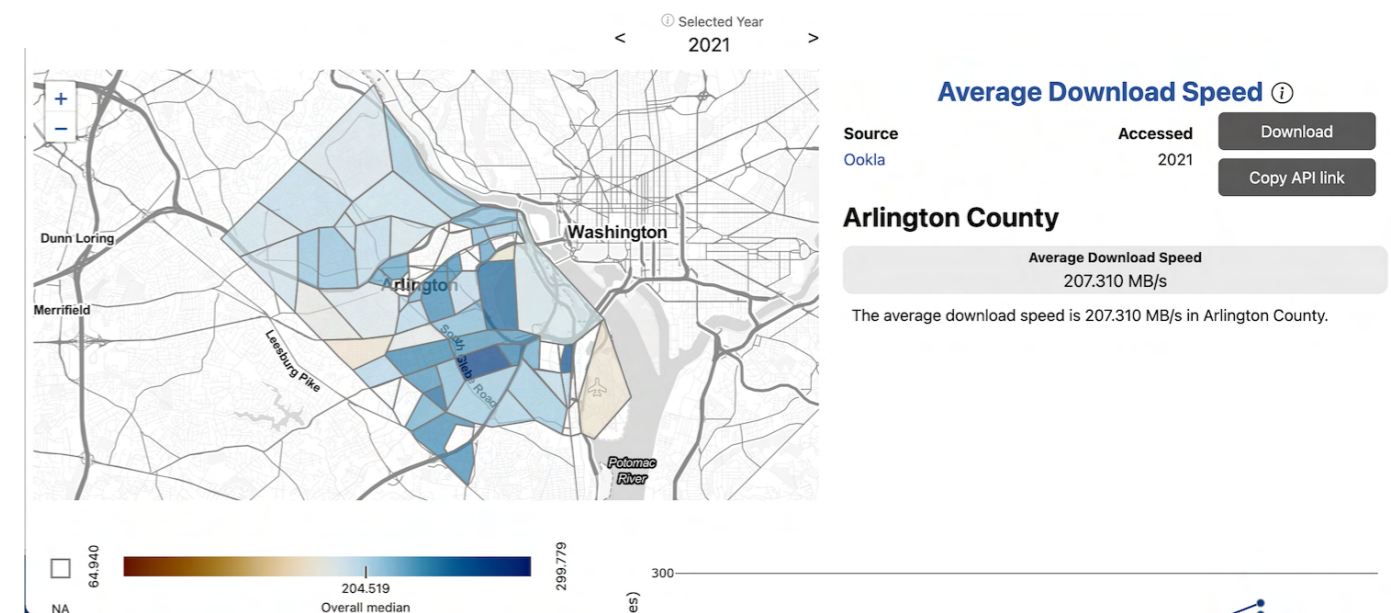
“We created local data insights and visualizations illustrating whether broadband adoption is equitably distributed in multiple localities in the Washington, D.C., metro areas,” Schroeder said. “What we can see from the data is that the areas with the lowest broadband adoption appear to directly correlate with the areas having the highest ratio of household income to the cost of broadband, indicating that broadband use is an economic issue as opposed to one of availability. This data was presented in local city council meetings and is directly informing the region’s resource planning.”

The team is developing the Social Impact Data Commons with transparent technology that is replicable and scalable for other regions undergoing economic change. All developed data, code, and tools are maintained as open-source repositories.

Co-Principal Investigator, Social and Decision Analytics Division Director, and Distinguished Professor in Biocomplexity Sallie Keller said, **“We built the Data Commons to be scaled and replicated to democratize data through data insights for local communities across the country. If we can achieve that, then we are living up to UVA’s vision to be ‘both great and good’ in all that we do.”**

LEARN MORE: SOCIAL IMPACT DATA COMMONS MICROSITE

## ARLINGTON BROADBAND INFRASTRUCTURE





*CONNECTIONS TO*

**DRIVE INNOVATION**



# Digital Twin Technology: The Future is Here

## Biocomplexity Institute Pioneers Novel Technology for Synthetic Populations

From contagions to natural disasters, power grid planning to elections, researchers are increasingly relying on sophisticated computational models to improve planning, response, and decision making for a wide range of scenarios.

Using high-performance computing, the Biocomplexity Institute's Network Systems Science and Advanced Computing (NSSAC) division can develop **highly accurate synthetic replicas or "digital twins"** of population sets throughout the world. This data has unlocked the ability for officials at the local level to the White House to **better understand and respond** to some of the world's most complex problems.

The Institute's digital twin technology captures detailed population data at both the macro and micro levels, including basic demographic properties, household structure, activity schedules, locations people visit, and where they interact with others. While extremely detailed, the data is carefully culled to preserve the privacy of individuals and protect sensitive data. Researchers across disciplines are rapidly adopting synthetic data to simulate "what if" scenarios in situations where there is an absence of real data or where maintaining privacy is integral.

## Diverse Application

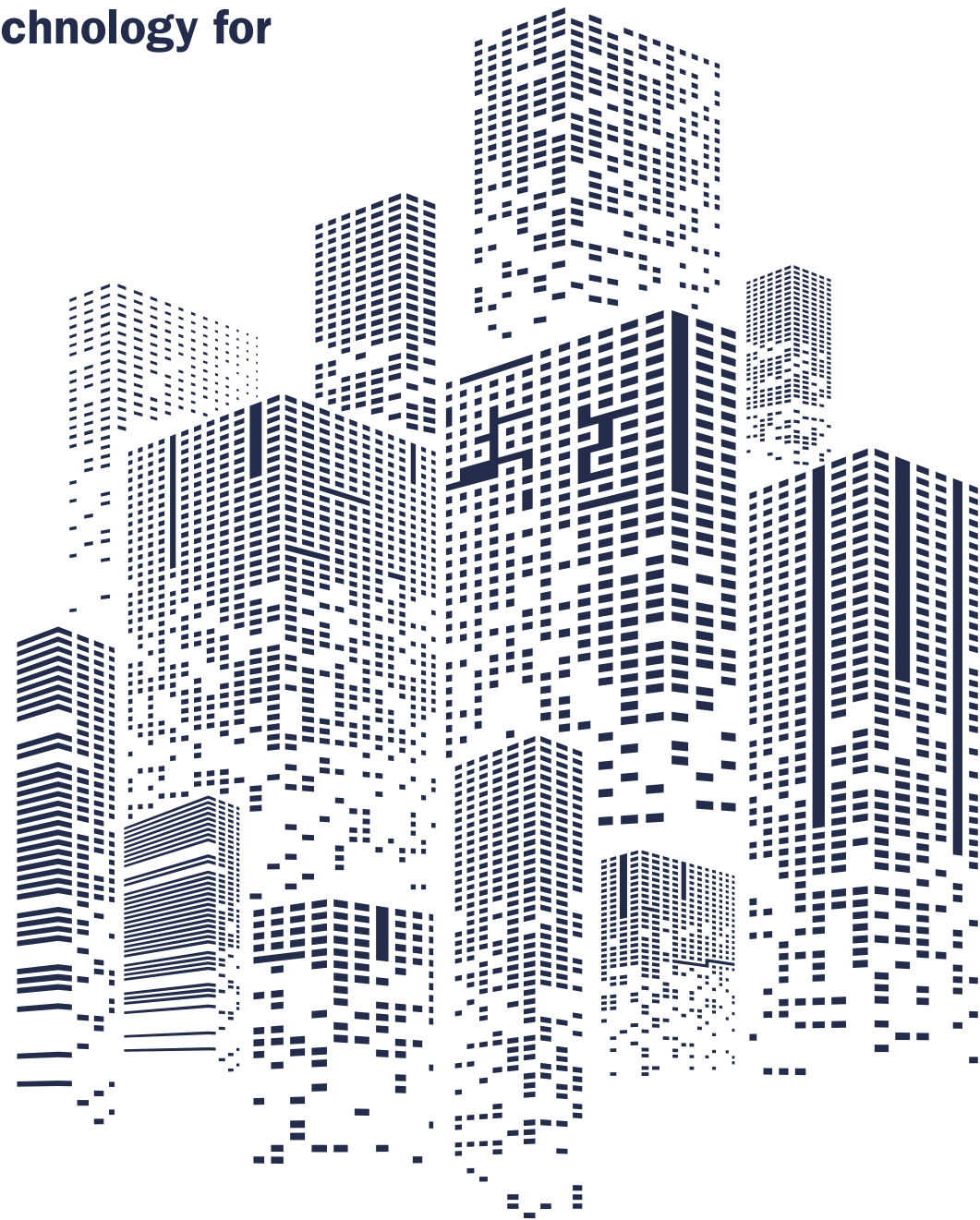
NSSAC division researchers have developed modeling frameworks using digital twins for a variety of applications including:

### Energy Consumption In U.S. Households

To achieve sustainable energy goals in the era of climate change and grid modernization, efficient energy consumption is essential. Therefore, we must understand how energy is consumed at a granular, household level to analyze the impact of factors such as weather, solar, electricity prices, electric vehicles, and occupancy schedules. In the absence of relevant data to study the modern electric grid, Institute researchers generated a large-scale, synthetic model of residential energy use for millions of households in the United States to help determine how to adapt energy infrastructure to meet the challenges of climate change.

### Ensuring Network Reliability Amidst Rising Use of Electric Vehicles

As more people move to electric vehicles (EV), consumers have become active participants in the power distribution network when choosing when to charge them. This creates a challenge for network operators in dispatching reliable power without taxing grid capacity. Institute researchers employed a digital twin of an actual power distribution network to validate a proposed EV charging schedule based on consumer preferences with respect to network constraints. Models revealed that the new approach would enable the existing network to keep up with increasing EV adoption rates without significant investments in grid infrastructure.



## NOVEL TECHNOLOGIES

### COVID-19 Forecasting to Project Outbreaks

Using a digital twin of the U.S. population, the Institute team developed a framework to generate real-time scenario projections to forecast outbreak hot spots and severity at the county level for the entire country on a weekly basis. Using multi-scale models, the Institute's NSSAC team contributed outbreak forecasting to national COVID response teams and senior leadership at the Centers for Disease Control and Prevention (CDC). This work garnered national acclaim as the Institute was named as one of five finalists for the ACM Gordon Bell Special Prize for High Performance Computing-Based COVID-19 Research.

### Institute Supports White House Challenge

In December 2021, the United States and United Kingdom announced plans for a collaborative \$1.6 million prize challenge to advance privacy-enhancing technologies (PETs). The competition aims to harness the power of PETs in solving two specific challenges – improving detection of financial crime and bolstering pandemic response capabilities by forecasting an individual's risk of infection. For the COVID-related track, the Biocomplexity Institute provided competitors with access to a digital twin of a regional population.

According to the White House, "This emerging group of technologies present an important opportunity to harness the power of data in a manner that protects privacy and intellectual property, enabling cross-border and cross-sector collaboration to solve shared challenges." Winners will be announced at the Summit for Democracy in the first half of 2023.

**"We are on the cusp of solving some of the world's most intractable problems and improving our quality of life with the power of artificial intelligence, but we must do it responsibly by upholding our shared values around privacy,"** said U.S. Secretary of Commerce Gina Raimondo. **"I'm thrilled that we're launching these joint U.K.-U.S. privacy-enhancing technology prize challenges and motivating our best researchers in industry and academia to innovate on protecting privacy so that we can all reap the benefits."**

## NOVEL TECHNOLOGIES

**"Digital twin technology is revolutionizing the way we plan and respond to our evolving world and its emerging challenges,"** NSSAC Division Director Dr. Madhav Marathe said. **"With data integrity and privacy paramount, the potential application for digital twins throughout academia, business, and government and across disciplines is limitless.** The ongoing White House challenge demonstrates how such digital twins can be used in **important national initiatives,** and we are honored to be at the **forefront of this technology."**

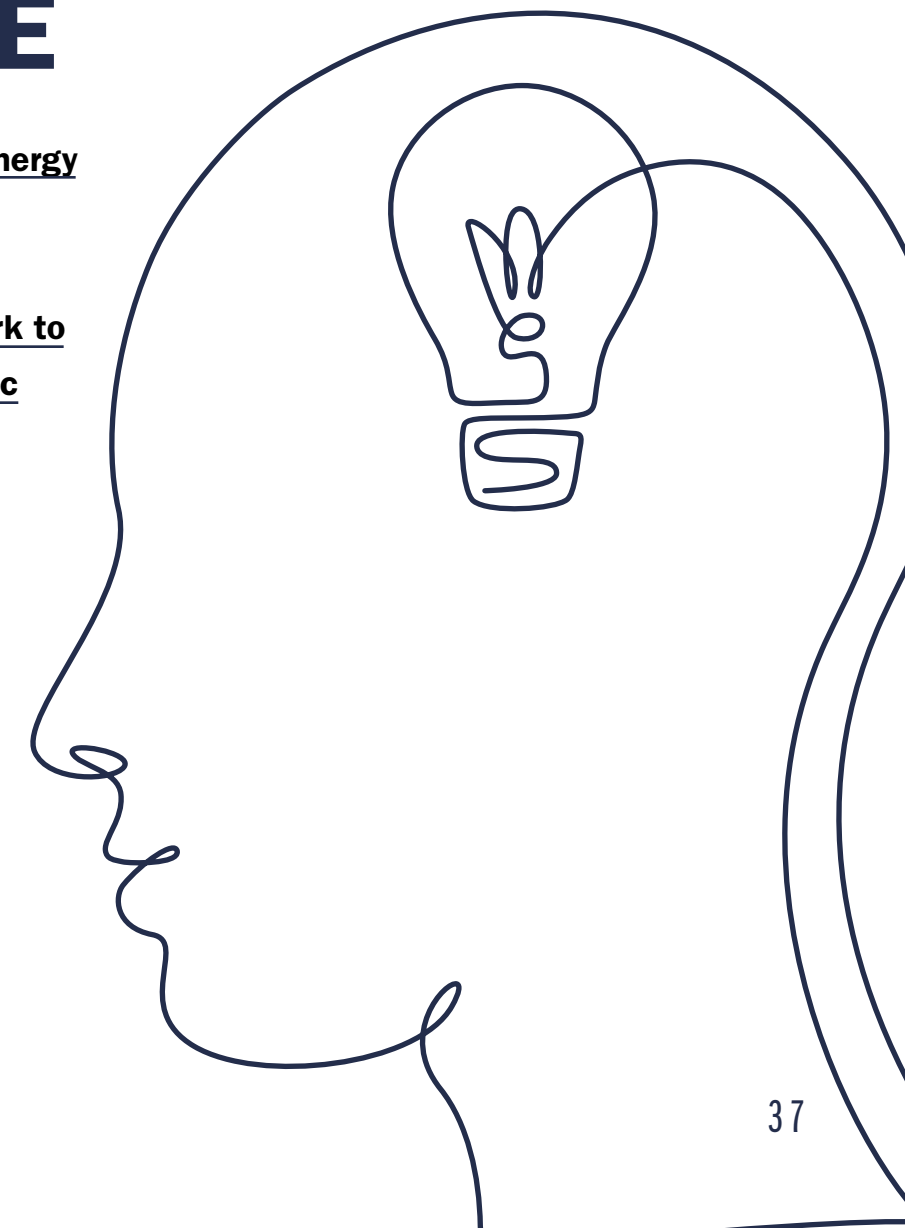
## LEARN MORE

[High Resolution Synthetic Residential Energy Use Profiles for the United States](#)

[A Reliability-aware Distributed Framework to Schedule Residential Charging of Electric Vehicles](#)

[PRESS RELEASE: U.S. and U.K. Launch Innovation Prize Challenges in Privacy-Enhancing Technologies to Tackle Financial Crime and Public Health Emergencies](#)

[Biocomplexity Institute's COVID-19 Pandemic Response](#)





CONNECTIONS FOR A  
**BRIGHTER FUTURE**

# DSPG Young Scholars Equips Tomorrow's Scientists to "Do Data Science"



The Social and Decision Analytics Division (SDAD) Data Science for the Public Good (DSPG) Young Scholars program brings undergraduate and graduate students from across the country together to engage with research projects that address local, state, and federal government challenges around critical social issues. The program equips a new generation of scientists with the skills to inform public policy and decision making around complex problems that impact people's lives in meaningful ways.

DSPG Young Scholars conduct research at the intersection of statistics, computer science, and the social sciences, such as economics, sociology, and psychology, to determine how information generated within communities and through government and community programs can be leveraged to improve quality of life. DSPG Young Scholars work in collaborative, vertically and horizontally integrated teams alongside postdoctoral associates, research faculty, and support staff from SDAD. Hallmarks of the program include extensive training through workshops, seminars, and career talks, individualized mentorship by division researchers, and unique opportunities to interact with decision makers from across community groups, industry, and government agencies.

In 2022, the DSPG program returned to an in-person format, having functioned as a virtual program for the past two years. Seven Young Scholars with backgrounds in computer engineering, biostatistics, economics, computer science, mathematics, and statistics joined SDAD in its Rosslyn, Va. office from May 30 to August 5. Each scholar supported two projects from among six different ongoing research collaborations, including:

- Constructing two data commons, one focused on social equity across the National Capital Region (sponsored by the Mastercard Center for Inclusive Growth) and one addressing inequality in health outcomes between urban and rural areas in Virginia (sponsored by the Virginia Department of Health)
- Building capacity for data-driven adaptation in rural Virginia coastal communities in response to changing climate conditions in conjunction with partners from across UVA
- Evaluating the impacts of broadband development on rural property values for the U.S. Department of Agriculture



- Conducting a qualitative analysis of U.S. Army and academic literature to identify individual characteristics in Soldiers that contribute to unit performance for the U.S. Army Research Institute for the Behavioral and Social Sciences
- Building a proof-of-concept toolkit to sustainably track innovation activities using opportunity data for the National Center for Science and Engineering Statistics (NCSES)

## EXPERIENTIAL LEARNING

- Identifying and tracking emerging digitalization trends in federally funded research and development grants for NCSES

The 2022 DSPG program also featured a variety of guest speakers from Microsoft, the National Science Foundation, and the Urban Institute, among others. They addressed topics such as the challenges of protecting personal information in publicly available data, using data science to generate innovative ideas for addressing gender inequality, and predicting and preventing extremist violence.

The Biocomplexity Institute and Initiative (BII) also built upon its longtime partnership with Virginia Tech, which organizes a sister DSPG program, hosting the visiting scholars at the BII's Northern Virginia location for an afternoon of research sharing and socializing. The DSPG Young Scholars also connected with participants in the UVA School of Data Science's Data Justice Academy, hosting them for a presentation from Aeyon and making a site visit to Deloitte Consulting to hear about several of Deloitte's data science initiatives.

Capping off the program, the 2022 Young Scholars shared the output of their research at the annual DSPG Symposium on August 4. In addition to the student presentations, the symposium featured a keynote address by Joseph Salvo, former chief demographer at the New York City Department of City Planning, who spoke about "Better Data and Measurement for Local Public Policy Decisions."



## EXPERIENTIAL LEARNING

# Computing for Global Challenges: Training the Next Generation to Tackle Our World's Most Pivotal Issues

Each summer, the Biocomplexity Institute and Initiative's Network Systems Science and Advanced Computing (NSSAC) division presents the Computing for Global Challenges (C4GC) program – a hands-on, immersive, eight-week training ground for the next generation of data-driven researchers. Working in teams with peers and NSSAC scientists, undergraduate students learn about cutting-edge software technologies and methods in machine learning, network science, simulation, data science, and computational biology and how these methods can be applied to address pressing real-world issues from climate change to epidemic forecasting and simulations.

The primary goal of the C4GC program is to provide a high-quality education for the students that they cannot easily obtain in a traditional classroom setting through: 1) hands-on exposure to research science; and, 2) experience working in a professional academic environment. With these targets in mind, the program aims for students to meaningfully contribute to their research assignments, which are sourced from the Institute's pool of active funded projects. As part of this, the interns are encouraged to develop close working relationships with their mentors and structure the program around that relationship and workflow.

The C4GC program endeavors to create an environment in which interns may make material contributions to their project, develop lasting connections with other students and mentors, and improve common professional skills, such as problem solving, communication, and resourcefulness. The 2022 program hosted various events in pursuit of these goals in addition to the day-to-day work between the interns and their mentors. The interns presented their work weekly, providing an opportunity to practice presentation skills and for their peers to gain a broader understanding of the Institute's work and scope.



**“I appreciated the intern meetings where we could see what projects everyone else was working on,” said Allison Lai, a third-year student in the UVA Computer Science program. “There were so many **different topics** and I like that there was a good mix from **algorithm-focused** to **visualization-based**. As the majority of us were working on projects alone, it can be **inspiring** to see other ideas, libraries, etc. that students used and trying it out myself.”**

## EXPERIENTIAL LEARNING

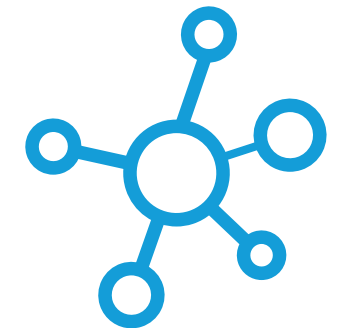
This year, the Institute expanded the program to accept 30 students – a 67 percent increase over 2021 – paired with 18 mentors. The program’s growth was the result of both a larger and stronger applicant pool as well as an increased number of qualified Institute projects.

The 2022 class was diverse along both the scholastic and personal spectrums, including: a range of first-year to fourth-year students; majors including mathematics, statistics, computer science, cognitive science, economics, global public health, and psychology; and, 20 percent of participants identifying as women. The program also accepted seven students from other institutions and three at the high-school level.

The program concluded at the end of the summer with its annual symposium where the students presented their work on more than 20 different research projects. According to post-program surveys, more than 90 percent of 2022 participants believed that the program would be relevant or extremely relevant to their career aspirations and over 95 percent would recommend the program to others.

**“Overall, the C4CG program was a really valuable experience,”** said Nicholas Wu, a fourth-year student in the UVA Engineering Science program. **“I hope more students can continue to learn at the Institute in future summers.”**

## LEARN MORE



[2022 RESEARCH PROJECTS](#)



[COMPUTING FOR GLOBAL CHALLENGES  
WEBPAGE](#)

## HOW TO GIVE

The Biocomplexity Institute and Initiative (BII) at the University of Virginia combines multi-disciplinary research expertise, advanced mathematical analysis, and extreme scale computing to understand and improve the health, habitat, and well-being of people. Our team science approach uniquely positions us to provide government, industry, and other stakeholders with the information and technologies to anticipate and respond to complicated challenges.

We need your support to confront the problems of today and those emerging tomorrow, including our work to:

- Model strategies to promote resiliency in the face of natural disasters;
- Assess the economic and educational impact of broadband access in rural communities;
- Combat COVID-19, malaria, Zika, and other vector-borne diseases and contagions around the globe;
- Redress national-scale critical infrastructure challenges like public transportation, air pollution, sea level rise, and supply chain crises;
- Develop a new and scalable data commons to optimize the next US Census; and
- Support US national security.

BII provides unique opportunities for corporate partners to engage with us in translational research, including in government and civilian contracting opportunities—and maintains a seasoned pre- and post-award development team to support customers in the public and private sectors.

**Please donate to support BII's work, or contact us directly to explore opportunities with us.**

[DONATE HERE](#)





---

## ACKNOWLEDGEMENTS

# Acknowledgements

We, the senior leaders of the Biocomplexity Institute, wish to recognize and applaud the creation of foundational knowledge and applied solutions to dynamic, real-world challenges through the concerted effort of all faculty, staff, and students employed by the Institute as organized in its four divisions: Network Systems Science and Advanced Computing, Social and Decision Analytics, Mathematical Biocomplexity, and Research Operations (Administrative). A framework of thematic research and transdisciplinary respect throughout the Institute underpins all scientific and operational activities and allows the research community of the Institute to thrive.

We offer a special note of gratitude to our distinguished faculty and scholars. Your mentorship and guidance transform new generations of scientists, trainees, and students in our care.

### Executive Director

#### Chris Barrett

Distinguished Professor in Biocomplexity  
Professor of Computer Science, School of Engineering and Applied Science

### Division Directors

#### Lauren Coble

Administrative Division  
Chief Operating Officer

#### Stephanie Shipp

Social and Decision Analytics Division  
Professor and Interim Division Director

#### Madhav Marathe

Network Systems Science and Advanced Computing Division  
Distinguished Professor in Biocomplexity  
Professor of Computer Science, School of Engineering and Applied Science

#### Christian Reidys

Mathematical Biocomplexity Division  
Professor of Mathematics, College of Arts and Sciences

#### Kevin Passarello

Director, Translational Programs and Development

For the FY 2022 Annual Report, we extend special thanks to Jill Draughon for the graphic design and layout, to Stacey Sepp, Xstatic PR, for framing our scientific themes and achievements, and to Grace Dusseau for the conceptual design.

For more information, go to our [website](#) or view the [Biocomplexity Institute Appendices for 2022](#).



---

**BIOCOMPLEXITY INSTITUTE**