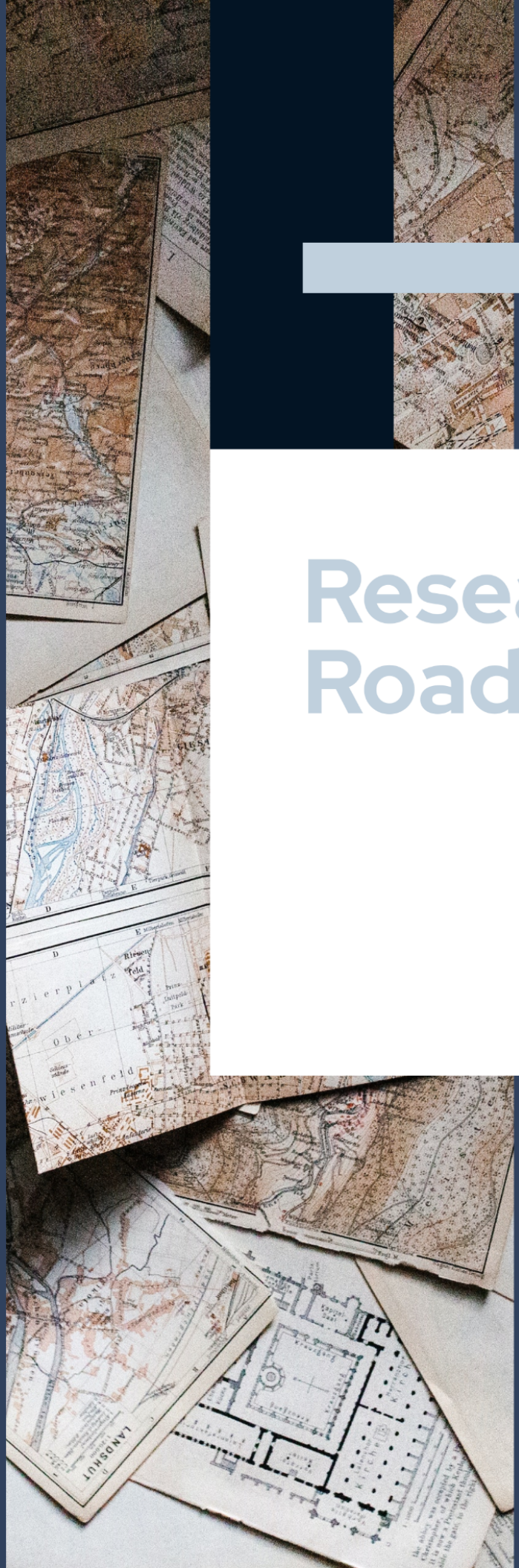


PREPARE

Pandemic Research for Preparedness & Resilience

Research Roadmap

Madhav Marathe, UVA
Simon Levin, Princeton Univ
Anil Vullikanti, UVA
Li Xiong, Emory Univ
Erin Raymond, UVA
Golda Barrow, UVA
Lily Li, UVA

September 2023 Report

Acknowledgements

The team would like to thank NSF for this opportunity to interact with our research community in such an impactful way. We have benefitted more than we can express, and we hope that the results of the PREPARE project can truly lead to a better prepared and more resilient global community ready to join forces in the battle against the next pandemic. Not with hubris or myopic overconfidence, but with the certainty that this virtual community will stand together with unflinching determination to provide innovative solutions toward the most effective response possible.

We would also like to recognize the contributions of researchers and scientists around the world who are working everyday to improve public health. From classrooms to field labs, hospitals to data centers - we see you, and we are humbly indebted to your contributions.

This material is based upon work supported by the National Science Foundation under Grant No. CNS-2041952.

NSF PREPARE Research Roadmap

EXECUTIVE SUMMARY

PREPARE (Pandemic Research for Preparedness and Resilience) is an NSF CISE-sponsored virtual organization tasked with fostering research collaborations and synthesizing critical pandemic-related computing research into a roadmap to help inform NSF funding opportunities that will aid our nation's effective response to the next pandemic. Since we started this project in October 2020, we have hosted eight virtual workshops featuring 72 subject-matter experts as speakers, panelists, and committee members. Collectively, these sessions were attended by over 2000 researchers and viewed on YouTube more than 3800 times. Please see prepare-vo.org for more details.

Through the aforementioned workshops, plus conversations with community members, podcast interviews, and literature review, we have gathered a good deal of information which we have synthesized as input into a set of recommendations meant to advise NSF leadership as they determine funding for programs that will help our world prepare to take on the next pandemic. This work represents input from a multidisciplinary assemblage of international researchers, and recommendations are offered in the following areas: Importance of Multidisciplinary Collaborations and Industry-Academia-Government (IAG) Partnerships; Cyberinfrastructure, Data, Data Analysis, and Responsible AI and Tools; and Societal Impacts. The appendices include an index of the recommendations, statistics about our events, summaries of each workshop with links to the original content, and a non-exclusive list of content contributors.

In the very wide spectrum of pandemic response, the following recommendations are specifically focused on the NSF CISE community, although the multidisciplinary nature of the research response to an event of the magnitude of COVID-19 cannot be ignored. The intention of the PREPARE team is that the Research Roadmap can act as an effective tool for NSF leadership, and significantly benefit international research initiatives focused on pandemic resilience and preparedness.

I. OVERVIEW

As COVID-19 ravaged the world during the summer of 2020, it became abundantly evident that a concerted effort to unify disparate threads of pandemic research funded through the NSF RAPID program would be an essential tool in mitigating and eventually overcoming the damage wrought by the pandemic. NSF leadership determined that a virtual community of researchers under the umbrella of a single organization could ideally distill this research into a single roadmap-style document to inform the agency about gaps, challenges, and opportunities for future funded thematic lines. The virtual organization PREPARE (Pandemic Research for Resilience and Preparedness)[1] was created to facilitate communication and collaboration among NSF CISE-funded scientists involved in pandemic research. With guidance from a Steering Committee composed of members from industry, academia, and government agencies, this organization enables the sharing of research results in a way only possible with a concerted effort. The depth and breadth of multidisciplinary collaboration enabled by PREPARE incorporates researchers, educators, and students interested in general topics related to pandemic planning and resilience. Utilizing a variety of dissemination platforms, PREPARE aims to harness the synergies of a broad range of research programs to facilitate scientific developments and advance public health emergency response in the US and around the world.

The increased levels of collaboration fostered by PREPARE have the potential to result in innovative computational methods and technologies for addressing future pandemics. The Research Roadmap includes the identification of key research topics, risks, and gaps in the current R&D landscape that when considered may significantly benefit the research community and serve as a blueprint for researchers, funding agencies, and policy makers on the role of information and communication technologies (ICT) in developing break-through solutions for pandemic resilience. While we are unlikely to prevent pandemics completely, developing resilience techniques will enable society to prepare for and cope with the aftermath in a more effective manner, specifically aided by NSF-funded research programs.

This Research Roadmap is a synthesis of the team's interactions with a variety of stakeholders immersed in pandemic preparedness and public health emergency response. We have met formally with our Steering Committee four times since its formation and have engaged with members through other channels on multiple occasions. Our workshops have included discussions around: the access, creation, and maintenance of data and computing resources; social, behavioral, economic, and governance issues during a pandemic; scalable computing, vaccine preventable diseases in a post-COVID world; and lessons and experiences on viable epidemic response strategies. We have worked with 72

[1] https://prepare-vo.org/

subject-matter experts as speakers, panelists, and organizers of these workshops. Our total virtual attendance number among all sessions is over 2000 people, with an additional 3800+ views on YouTube[2]. We continue to engage with several participants after these workshops, keeping the dialogue alive within the community. We also maintain interaction with community members through our *Science Before the Storm* podcast[3], where we have enjoyed conversations with several scientists over our three seasons of episodes. And finally, we have looked at a variety of reports and publications to further inform our recommendations[4].

The research community is still very invested in the topic of pandemic preparedness, and the high attendance at our January 2023 workshop highlights this interest level. Our LEVERS (*Lessons and Experiences on Viable Epidemic Response Strategies*) workshop had 401 registrations, and has been viewed on YouTube 860+ times as of September 2023. We believe that NSF PREPARE is the perfect vehicle through which the scientists can continue to interact on a multidisciplinary international level, bringing the results of their cutting-edge research to a community eager to establish tools and resources that can be quickly deployed when the next pandemic strikes.

PREPARE was initially charged to support the NSF CISE directorate's funded domestic RAPID projects. However, as we met with stakeholders in this community, it became progressively clearer that we needed to broaden the scope of our efforts. We thus began to incorporate research from all NSF directorates to gain a more robust and accurate understanding of the gaps and challenges we face in pandemic preparedness and resilience. It has also become apparent that in the globally mobile world of modern times, any pandemic response efforts must integrate international partners. We thus began including panelists and program committee members from countries like India, Denmark, Australia, Hong Kong, Rwanda, and the UK to build bridges beyond our domestic borders. Our interactions also brought to fore the need for synthetic data for technology development, training AI methods, and developing explainable systems, and our team's recent involvement in the UK-US Privacy Enhancing Technologies Prize Challenge underscores this need (see https://petsprizechallenges.com/).

The PREPARE team condensed and synthesized this material into a series of recommendations for NSF to consider in support of future research directions. This work represents input from a truly multidisciplinary assemblage of international researchers and is broken into the following areas: Importance of Multidisciplinary Collaborations and Industry-Academia-Government (IAG)

---

[2] https://www.youtube.com/@nsf_prepare

[3] https://open.spotify.com/show/49HJAQ6UxUIZM7shzbpA7V?si=f18d322811854780

[4] Given the immense number of publications on various topics related to the pandemics, we have chosen a few reports and summary articles that capture key recommendations and insights.
https://theindependentpanel.org/documents/; https://www.nature.com/collections/jaacfgeief;
ttps://www.mckinsey.com/featured-insights/themes/how-to-prepare-for-the-next-pandemic;
https://www.cell.com/cell-reports-medicine/pdf/S2666-3791(22)00431-1.pdf;
https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf?adobe_mc=MCORGID%3D242B6472541199F70A4C98A6%2540AdobeOrg%7CTS%3D1681960891;
https://www.nascio.org/wp-content/uploads/2021/10/NASCIO-CDG-IBM-AI-Meets-the-Moment-2021.pdf;
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1128099/Technical-report-on-the-COVID-19-pandemic-in-the-UK-PRINT.pdf; https://science.osti.gov/-/media/Initiatives/pdf/Bioprepardeness_Roundtable_Report_092722.pdf;
https://www.nature.com/collections/ibajaahdgh; https://www.nature.com/articles/s41586-019-1717-y;
https://www.hhs.gov/sites/default/files/draft-paccarb-pandemic-preparedness-report.pdf;
https://www.whitehouse.gov/wp-content/uploads/2021/09/American-Pandemic-Preparedness-Transforming-Our-Capabilities-Final-For-Web.pdf?page=29;
https://www.whitehouse.gov/wp-content/uploads/2021/01/National-Strategy-for-the-COVID-19-Response-and-Pandemic-Preparedness.pdf;
https://www.cdc.gov/nchs/data/misc/STPI-Epi-Modeling-Summit-Report-01-14-2021.pdf

Partnerships; Cyberinfrastructure, Data, Data Analysis, and Responsible AI and Tools; and Societal Impacts. We have also included holistic recommendations for NSF and ideas for future discussions to collect additional input for the Research Roadmap. The appendices include an index of the recommendations, statistics about our events, summaries of each workshop with links to the original content, and a non-exclusive list of content contributors.

Note that these recommendations are focused around NSF-funded research, particularly in the area of computing. In a very broad sense, the centrality of computing in effective global pandemic response efforts reinforces the solid position of the NSF CISE Directorate at the intersection of the multidisciplinary response required to combat the next pandemic[5].

Some of the PREPARE committee members, speakers, panelists, and podcast guests. For a more comprehensive list, see Appendix E.

[5] Alghamdi N, Alghamdi SM. The Role of Digital Technology in Curbing COVID-19. 2022 Jul 7 [cited 2023 Jul 21];19(14):8287–7. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9320375/

II. Importance of Multidisciplinary Collaboration and Industry-Academia-Government (IAG) Partnerships

As the PREPARE team interacted with colleagues from around the world, a common recommendation included the absolute necessity for a framework that encouraged and supported interactions across disciplines with a wide breadth of stakeholders. One major barrier to successful pandemic preparedness was the lack of easily identifiable bridges to connect researchers to each other and to the public health decision makers who may have felt better equipped to address the wide-ranging issues they faced with the support of subject matter experts in academia. Our decision to host the January 2023 LEVERS (*Lessons and Experiences on Viable Epidemic Response Strategies*) workshop was motivated in part by the urgency of this need across sectors, disciplines, and nations.

Policy Makers and Public Health Officials

Recommendation 1
Organize and/or fund initiatives that bring together policy makers, public health officials, and researchers from diverse disciplines with the primary objectives of building rapport and opening dialogue to build collaborative relationships.

A key component of successful pandemic response is to ensure that policy makers at local, state, and federal levels are involved with researchers from the earliest stages of a public health crisis. Furthermore, it is necessary to have an integrated plan before the pandemic occurs and the plan needs to be tested frequently, communicated broadly, and adapted as conditions change. Consistent and significant interaction with policy makers is essential for widespread adoption of science-based public health recommendations, and those relationships must be established before the crisis. Communicating uncertainty is critical to informed decision making.

One example of a successful IAG collaboration program is the Emory COVID-19 Response Collaborative[6], in which public health faculty and students work with public health agencies through a two-year fellowship program. During this time, student fellows are assigned to one of the eighteen Georgia public health districts or the Georgia Department of Public Health. With interactions fostered by intentional IAG collaborations, researchers can study questions that would be otherwise very challenging. For example:

- What information do they wish they'd had when the pandemic started?
- How would they have handled the situation differently with that knowledge?
- How do we find a balance between public health, social cohesion, and economics?

Deciding collectively what data is most useful will help the research community to serve policy makers most effectively as end users. Discussing proposed interventions with relevant authorities may allow modelers to redirect or narrow efforts toward modeling a broad spectrum of implementable interventions. From this spectrum, policy makers may be able to implement light-touch interventions that involve more decentralized decision making (e.g., ways for people to assess the appropriate actions to take based on their specific situation) hopefully encouraging greater buy-in and thus compliance.

[6] https://emorycovidcollaborative.org/

National Agencies, Corporations, and Foundations

Recommendation 2
Develop coordination between US research funding agencies to ensure we have the capability to rapidly respond during a pandemic crisis.

There is strong interest in the community to have a national conversation on the need for pandemic preparedness among scientific agencies such as the National Academies, NSF, NIH, CDC, etc. Inter-agency partnerships should be a priority that lead to discovering, articulating, and publicly sharing the national (and global) agenda for pandemic preparedness research. Additionally, the work of foundations and corporations could be leveraged to build resources for future response needs.

One possible direction to consider is a targeted effort by NSF to fund focused research organizations (FROs)[7]. NSF has already demonstrated support of an FRO-type model through efforts like the May 2021 NSF Workshop on a National Network of Research Institutes (NNRI). We find our recommendations align well with those released in the workshop report[8].

Among the input we received, researchers made the following specific suggestions to address current challenges:

- Promote inter-agency funding for centers where the end users (policy makers/public health officials) are the PIs and there is a critical mass of experts from academia, industry, and government working together as a large-scale team to do groundbreaking work, translate CISE research into practice, and leverage advanced computing in industry to help accelerate scientific results; partnerships of this scope were specifically supported by participants at the NSF NNRI workshop as well

- Invest in emergency response infrastructure on a global level, from community engagement to high performance computing (HPC) to laboratory facilities to knowledge base, that includes a high-speed/low-barrier mechanism for sharing results and data

- Create standards for code and data sharing, appropriate mechanisms for crediting the authors and developers, and containerization that can help facilitate model evaluation, production, reproducibility, and operations

- Provide funding for peace time maintenance and potential transition into practice of the NSF RAPID-supported data collection and code development; lack of continued funding mechanisms often leads to the inaccessibility and degradation of these assets over time

[7] Marblestone A, Gamick A, Kalil T, Martin C, Cvitkovic M, Rodriques SG. Unblock research bottlenecks with non-profit start-ups. *Nature*. 2022 Jan 13;601(7892):188-90
[8] Focused Research Organizations: A New Model for Scientific Research - Federation of American Scientists [Internet]. Federation of American Scientists. 2023 [cited 2023 Jul 21]. https://fas.org/publication/focused-research-organizations-a-new-model-for-scientific-research/

Multidisciplinary Collaboration

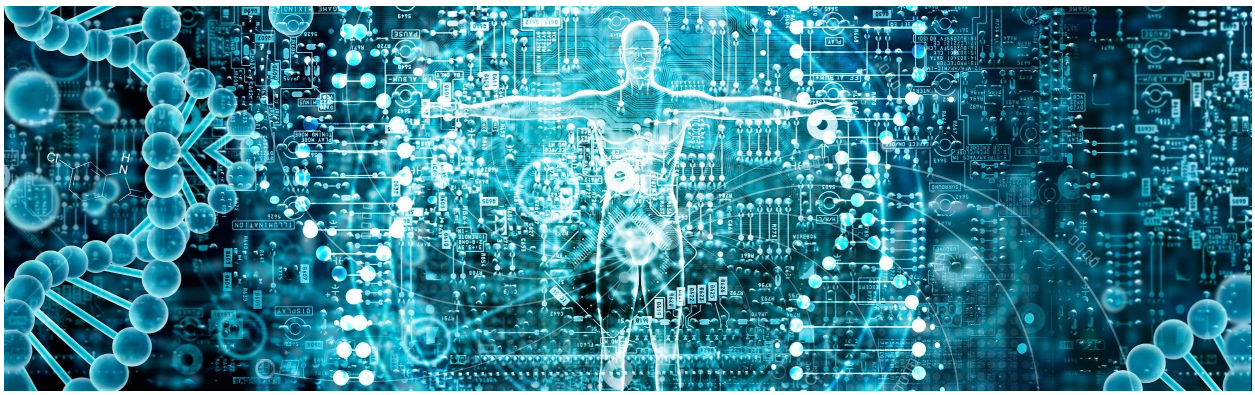
Recommendation 3
Advocate to increase the frequency and "typicality" of multidisciplinary research/collaboration at universities/research centers, funding agencies, and publication venues.

Many people we have engaged with have mentioned the essential nature of cross-functional collaboration, not just across disciplines but also across sectors, governments, and nations. Creating and maintaining a multidisciplinary infrastructure allows scientists to operate as one collective group more efficiently; these cross-disciplinary interactions are valuable and require resources. NSF can address these gaps with initiatives that:

- Encourage and fund longer term and scope multidisciplinary research; it takes time to develop a working relationship across traditional boundaries
- Ensure academic evaluation methods support multidisciplinary research

Researchers indicated that there are a spectrum of opportunities for collaboration, varying from videographers who can help communicate scientific results to the public to geneticists who can provide very specialized services. Supporting and advocating for cross-functional collaboration will foster a community of multidisciplinary researchers in reserve and ready to constructively pivot toward the next global challenge. NSF can specifically advocate for this at US academic institutions, particularly on behalf of early career researchers. Overcoming the challenge of finding a common multidisciplinary language will enhance the complex collaborations in pandemic science today that require different disciplines to effectively communicate and work together on transdisciplinary research directions.

The Summer 2023 NSF Predictive Intelligence for Pandemic Prevention (PIPP) call for center proposals[9] is an excellent example of intentional fostering of multidisciplinary collaborations, specifically proposing to fund centers of pandemic-related research composed of a coordinated and integrated team of diverse researchers who can leverage their work to produce innovative solutions in this space.



[9] https://new.nsf.gov/funding/opportunities/predictive-intelligence-pandemic-prevention-phase

III. Cyberinfrastructure, Data, Data Analysis, and Responsible AI and Tools

We have combined the aspects of cyberinfrastructure, data, data analysis, and responsible AI and tools in one section to better reflect the larger issue - the integration of these components into a productive workflow.

Cyberinfrastructure

While advanced computing initiatives are critical within the NSF CISE Directorate, we found that many researchers acknowledged the significant role computing environments play in public health emergency response. This was particularly evident in the discussions among panelists in our March 2022 *Scalable Computing for Pandemic Preparedness* workshop. There is a high level of interest in expanding and developing effective HPC pipelines and systems during peace times; thinking through use-case idiosyncrasies in advance makes deployment more fluent when the crisis occurs.

Recommendation 4
Encourage development of an advanced computing cyberinfrastructure that is user-focused and ubiquitous, removing barriers to use and ensuring availability of dedicated resources in response efforts.

During the COVID-19 response, researchers familiar with advanced computing struggled to find adequate resources, and in some cases found the barriers to adoption and inflexible resource allocation to be very high. Modern HPC users require service-oriented architectures that allow one to connect to different resources in novel ways, and modifying workflows to fit the constraints of the HPC center was an onerous burden. Some ways to improve HPC adoption by the research community include:

- Create the infrastructure ecosystem during peace time in anticipation of a public health emergency, providing a dedicated virtual container for development, debugging and runtime, and perhaps incentives to participate
- Remove perceived impediments to access and include funding for a dedicated customer support system
- Build software platforms that are focused on the research work rather than the low-level details of how to run codes on HPC resources
- Address the growing complexity in modern computational research, including the manipulation of more complex data types in various forms of standardization to make data computable and easily ingestible into new algorithms
- Educate emerging communities about advanced computing resources, what is available, and how it can apply to their research
- Resolve the challenges of migrating workflows between systems and building multi-system workflows

Data

Researchers acknowledge the benefits of access to an unprecedented amount of data during the COVID-19 pandemic, but the sheer variety of originating sources and multimodal types as well as lack of access to other desirable but privacy-sensitive data revealed challenging practical issues. It is one of the reasons we chose to focus on the *Access, Creation, and Maintenance of Data and Computing Resources* as our inaugural topical workshop in May 2021. Considering these issues, and in an effort to encourage a coordinated response to a future public health crisis, this recommendation could prove helpful in overcoming these challenges.

Recommendation 5
Create a global data and computing infrastructure via privacy enhancing technology to archive cross-institution/ agency multimodal data, and analytic tools that were critical in mitigating the effects of the COVID-19 pandemic, allowing for rapid mobilization of these resources in times of crisis.

Within such an infrastructure, researchers could build privacy-preserving and secure data collection, sharing, and analysis tools, and create pilot testbeds where researchers can evaluate privacy-enhancing technologies (PETs) as well as analytical tools. The infrastructure can provide tiered access with potentially different granularity and fidelity levels by combining PETs and governance and procedural control. The infrastructure must allow for next-generation multidisciplinary collaboration to enable real-time pandemic prediction of emergence and mitigation, and include a breadth of data types including, for example: socioeconomic features, mobility patterns, behavioral data, county-level NPI, vaccine allocation, confirmed case counts, and hospitalizations, clinical data, virological surveillance, and a nationwide longitudinal cohort study. The scientific community must better understand which data was missing from the COVID-19 analysis that could have had the biggest impact on mitigation.

At the inaugural Summit for Democracy in December 2021, leaders from the United Kingdom and United States announced a special privacy-enhancing technologies (PETs) prize challenge[10] focusing on pandemic risk prediction using decentralized privacy-sensitive mobility and contact data. PETs, such as federated learning, differential privacy, and homomorphic encryption, allow organizations to share and analyze sensitive data while protecting individual privacy. This enables the development of data-driven solutions to global challenges, while upholding democratic principles and fostering innovative and trustworthy applications. Members of the PREPARE team joined the University of Virginia Biocomplexity Institute's Network Systems Science and Advanced Computing division to create and provide a synthetic dataset for the challenge[11].

After the winners were recognized at a demonstration day in May 2023, the team participated in a roundtable event, "Supporting Pandemic Response and Healthcare with Privacy-Enhancing Technologies: Opportunities and Challenges."[12] The initial discussion focused on how data from various sources and modalities could be used to effectively support public health response during an evolving pandemic as well as general healthcare solutions. Participants then

[10] https://petsprizechallenges.com/
[11] https://prepare-vo.org/synthetic-pandemic-outbreaks
[12] NSF PREPARE. UK-US PETs Prize Challenge Roundtable Executive Summary. 2023 July. TR BI-2023-229.

discussed the application of privacy-enhancing technologies to unlock data to support public health and healthcare research, as well as pandemic response efforts. Participants identified broad challenges with privacy guarantees, identifying and accessing critical datasets, building an infrastructure to leverage the value of diverse datasets, and potential areas for future study and pilot solutions. They offered the following ideas for future directions:

- *Cost of privacy*. To gain a comprehensive understanding of the effectiveness of different pandemic response strategies implemented by countries during the SARS-CoV-2 pandemic, a comparative study across nations is needed. This study would assess case outcomes and determine the best approach. Evaluating the cost of privacy in terms of mortality (e.g. due to privacy barriers to contact tracing or lack of access to certain types of data) is crucial for making informed decisions.

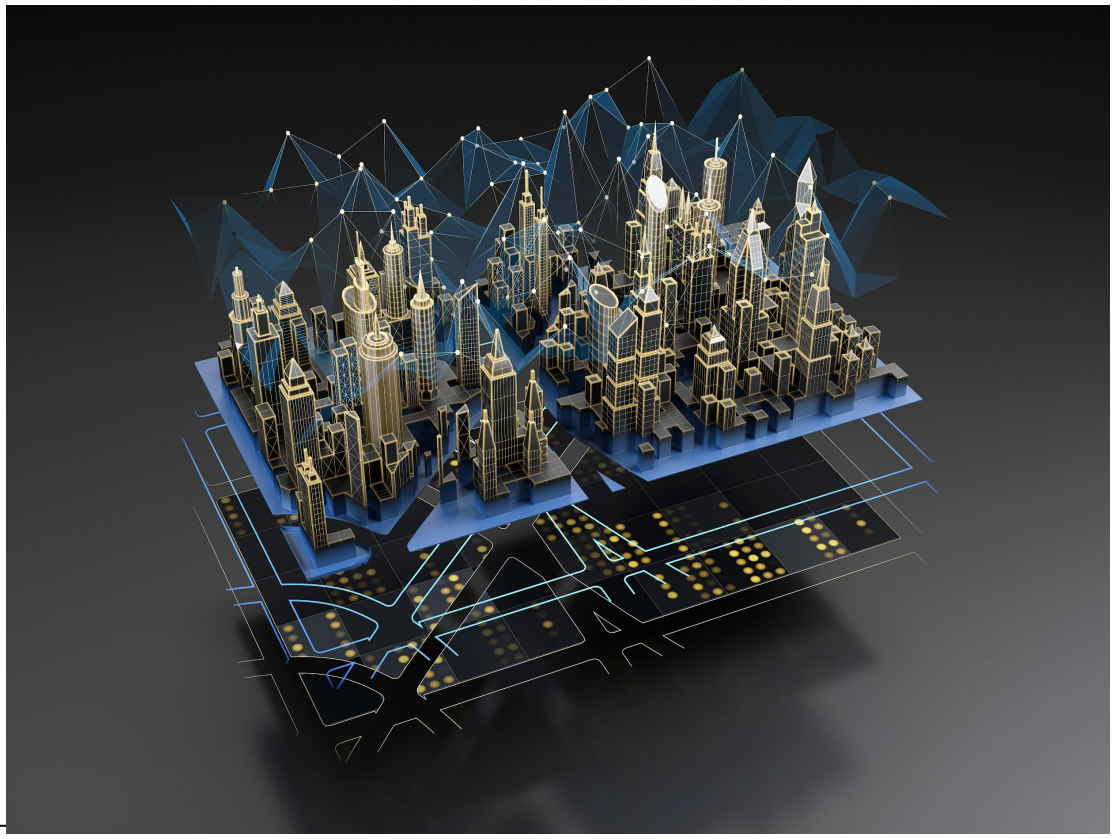
- *Leveraging mature technology*. Matching data use cases to the privacy risk/threat model is essential in determining the appropriate PETs to employ. For example, genomic privacy protection is critical particularly for rare genetic childhood diseases. Mature technology in certain areas can inform the development of PETs and systems in other health contexts. All the winning solutions of the PETs challenge utilized a combination of different PETs including federated learning, differential privacy and homomorphic encryption, to address different risks at different stages (e.g. data leakage risks during computation and inference risks from the learned models post computation), demonstrating the different use cases and tradeoffs offered by different PETs and the need to develop hybrid solutions and properly quantify the privacy guarantee of the composed solutions.

- *Data linkage.* Linking vertically partitioned patient data for a longitudinal view is critical but presents technical and systems research challenges. Additionally, the challenges posed by non-textual and multimodal data, as well as the integration of emerging technologies like telehealth, AR, and VR, must be addressed.

- *Systems and fundamental research.* Scaling the research requires collaboration, standardization, and interoperability across data owners and providers. Systems research, incentives, and regulatory structures play a vital role in this endeavor, and the relationship between systems integration research and fundamental research should be acknowledged to ensure a comprehensive approach.

Recommendation 5.1
Build a robust network of open synthetic data sources as an alternative to enhancing data sharing and access capabilities.

While many organizations willingly shared data during the earliest stages of the COVID-19 pandemic, as the global response became less urgent, those data resources became harder to access. For the owners of this data, there is a constant tension between serving the public good, preserving individual privacy, and navigating commercial incentives. One potential path through this roadblock is the creation of multiple substantial and diverse synthetic datasets using agent-based simulations or generative models. An example of such a dataset was recently created by members of the PREPARE team with additional researchers from UVA for the 2022 UK-US Prize Challenge on Privacy Enhancing Technologies. This synthetic data, sometimes called a virtual population or digital twin, is a statistically accurate representation of a real population's demographics, activities, and social contacts, but does not contain any individual person's information[13]. The challenge is how to ensure the data are realistic for different modeling scenarios and needs, capturing the complex patterns and dependencies in real data, representing the population without bias, while having formal privacy protection guarantees if private data sources are used during the synthesization process.



[13] Harrison G, Chen J, Mortveit H, Hoops S, Porebski P, Xie D, Wilson M, Bhattacharya P, Vullikanti A, Xiong L, Marathe M. Synthetic Data To Support US-UK Prize Challenge For Developing Privacy Enhancing Methods: Predicting Individual Infection Risk During A Pandemic [Data set]. doi:10.18130/V3/ZOG1FF

Data Analysis

Like navigating a highway system, pandemic research requires identifying key off-ramps that lead to important discoveries and breakthroughs. Researchers must be strategic in identifying and creating the tools and methods necessary to exploit these breakthroughs, increasing the collective capabilities at our disposal to recognize and mitigate public health threats. Two areas of focus here include computational modeling for foresight and biosurveillance, which were generally identified in our workshops as highly impactful in pandemic response.

Recommendation 6
Promote research that enhances and expands modeling capabilities, and support methods to communicate accurate interpretation to end users.

One tool that was front and center during the pandemic was modeling, although it is unclear if policy makers and the general public ever truly understood the nuances or correct application of these models. Focus should be placed on improving the quality of these models given the degree to which decision makers relied on them. One example of an ensemble modeling method that proved very useful to policy makers was the CDC-aligned COVID-19 Forecast Hub[14]. It was created with the goals of building a multi-model system that could provide reliable, short-term forecasts to the public and decision makers, assessing the reliability and success of different modeling approaches, and enabling the contributions of anyone interested in helping, established in the field or not[15]. The CDC-aligned COVID-19 Scenario Modeling Hub[16] addresses longer range questions which require capturing causal mechanisms at play.

Based on the successes and challenges of modeling efforts in response to COVID-19, researchers had the following broad suggestions:

- Support methodological advances to bridge AI/ML and develop multi-scale multi-theory models that take behavioral/causal, cognitive, political and economic, immunological and virological issues into account
- Develop generalizable computing and AI models that are scalable, explainable, and fair so as to be able to quickly adapt to changing scenarios and rapidly scale up
- Increase use of modeling to inform biosurveillance
- Create benchmarks to compare models
    - Predictive or retrospective validity of models, although useful, is not adequate
    - Model utility should be evaluated in terms of its ability to support effective decision making, situation assessment, and counterfactual analysis

[14] https://covid19forecasthub.org/
[15] Borchering, Rebecca K., et al. "Modeling of future COVID-19 cases, hospitalizations, and deaths, by vaccination rates and non pharmaceutical intervention scenarios — United States, April–September 2021." *Morbidity and Mortality Weekly Report* 70.19 (2021): 719.
[16] https://covid19scenariomodelinghub.org/

Biosurveillance

Recommendation 7
Advocate for the establishment of a global biosurveillance infrastructure in conjunction with relevant government agencies that supports an international sample collection mechanism with long-term funding and unrestricted data sharing.

Rapid response is critical in a public health emergency, and biosurveillance is a tool that can provide a warning signal for a potential danger ahead. Biosurveillance can aid public health officials in quickly identifying an emerging pathogen and deploying appropriate rapid response measures. Many participants in our June 2022 *Vaccine Preventable Diseases* workshop alluded to the need for this on an international scale. In a population with global mobility, international cooperation can help to mitigate the impact of pathogen spread.

The current system of national focal points of biosurveillance produces unacceptable delays in sharing along with single-point-of-failure choke points in both receiving and sending data. What's more, current efforts to tie "benefit sharing" to "data sharing" are likely to extend beyond biological samples to genomic sequence data. Some researchers believe the personal international collaborations within the scientific and biomedical communities – where they exist – have been much more effective than relying on government-level programs. How can we combine the stability of official channels with the effectiveness, timeliness, and robustness of personal communications? NSF can actively support international scientific collaboration, protect data sharing from government interference, and foster the wide dissemination of research results. To that end, NSF can support research to:

- Build a pathogen testing infrastructure that can detect the emergence of novel pathogens, similar to early warning systems for detecting earthquakes
- Create a more robust surveillance infrastructure, new surveillance apps, and effective tools for finding threats
- Connect and integrate such a biosurveillance system across scales from wastewater, to clinical, to societal aspects
- Develop technology and encourage the implementation of policy for efficient ways to do contact tracing and quarantine infrastructure
- Create an international structure of biosurveillance similar to what exists for influenza A where isolates can be collected from different countries, then sent to a series of accredited laboratories who have good sequencing facilities and technologies

Responsible AI and Tools

In an increasingly data-driven world, the responsible curation and use of data have become paramount concerns. Addressing issues of bias and fairness in data is not only a moral imperative but also essential for ensuring equitable access to valuable resources. To achieve this goal, it is crucial to support research endeavors that delve deep into the intricacies of data quality, uncovering the hidden biases and inequalities that often lurk within datasets. By fostering such research initiatives, we pave the way for the development of standardized practices that not only promote transparency and fairness but also empower individuals and organizations to harness the full potential of data while upholding ethical principles.

Recommendation 8
Promote responsible data curation and use through supporting research that addresses bias and fairness issues, ultimately creating standards for data quality and facilitating equitable resource access.

A common issue our community recognized is that there is inherent bias in data. For example, epidemiology data such as cases may be underreported in rural areas; digital epidemiology data such as search data, mobility data, social media data, and survey data, while providing complementary data signals, are limited to users of the platform. We need to understand the bias in the data when using the data to develop modeling and AI tools, including both the source and degree of bias or representativeness, and if possible, correct the bias to make data more representative of the underlying population, and ensure the analytical tools does not propagate and exaggerate the bias.

Another issue is that there is inherent noise, uncertainty, and inconsistency in data. For example, the reporting of daily cases vary significantly both temporally and regionally due to varying testing availability, reporting practices and standards. Additional uncertainty of the data may be also introduced by the privacy enhancing technology. Any use of this data needs to account for these factors.

To this end, NSF and other agencies can support research for:

- Appropriate quantification, calibration, verification, and external validation of data bias and quality

- Standardization of data collection and sharing
    - One example of an area where data standardization would be valuable is consistency in measuring and quantifying behavioral data. In all cases, legal and ethical issues for data sharing must be considered. As an agency, NSF may choose to promote potential legislation to enable data sharing for public health purposes, particularly in the case of a pandemic. Leveraging the relationships mentioned in Section II and promoting collaboration among companies, researchers, and government agencies could prove to be critical to ensuring global pandemic response preparedness. This might include providing incentives and resources for research groups and private companies to make their data FAIR (Findable,

Accessible, Interoperable, and Reusable), and funding mechanisms for facilitating, finding, and aggregating related data to open platforms for analysis.

- The interplay of privacy enhancing technology and data bias and fairness issues
  - For example, there is evidence that privacy-enhancing machine learning algorithms (e.g. with differential privacy) can amplify bias. Federated learning solutions can propagate bias from one site to others. More research is needed to develop responsible algorithms and tools that ensure privacy, fairness, and robustness simultaneously.

- Workforce education and training to ensure our next generation researchers and practitioners understand the nuances of data and algorithms and ensure responsible use of the data

IV. Societal Impacts

NSF PREPARE was funded through the visionary CISE Division, who quickly realized the need to capture and integrate the RAPID-funded pandemic response work into a roadmap that could inform future NSF CFPs to address gaps and challenges. We specifically selected *Social, Behavioral, Economic, and Governance Aspects of Pandemics* as the focus of our second topical workshop in June 2021 because during our interactions with researchers and policy makers, it became obvious that in our digital world the traditional CISE disciplines occupied the center of a heavily trafficked intersection.

Recommendation 9
Create cross-NSF directorate research initiatives focused on identifying and mitigating the spectrum of societal disparities and impacts exposed and aggravated as a result of COVID-19.

Researchers who attended our second annual meeting of pandemic researchers (RP2) listed clear and consistent scientific communication as the second greatest challenge during the pandemic - issues surrounding data imposed the greatest difficulties. This included: communication between researchers, who were possibly duplicating efforts due to the lack of available channels; communication between researchers and decision makers, where a lack of trust or knowledge of who to contact may have hampered interactions; and communication from researchers and decision makers to the general public, who were certainly overwhelmed by the many sources of digital information available. Some specific areas of study include:

- Understanding how people reacted to NPI and vaccination policies, including variability due to cultural and environmental factors and how those reactions changed over time, may generate insights for making future policy decisions and effective public messaging
- Understanding different socio-technical systems (for example, e-commerce, online shopping, online education) and how they impacted different household conditions
- Understanding healthcare disparities, including access to healthcare, surveillance system quality, and vaccine and therapeutics adoption
- Understanding the disparities of virtual working and stay-at-home orders, including inequitable access to high-speed internet and places for isolation and effective remote work, and the psychological impact on social interactions within communities to inform/leverage future technology development

- Advancing the scientific understanding of the polarization of pandemic discourse and overall information consumption, e.g., the ways social media are used, how the information is understood, the role of mis- and disinformation in decision making
- Analyzing the entire education system, including long-term studies to understand the impact of the pandemic and validate effective education delivery and educational infrastructure, and applying those lessons learned beyond the classroom to the workplace

VI. General Recommendations for NSF

During our RP2 meeting, we asked each presenter for their suggestions or requests for NSF. While we have incorporated many of these into the previous sections, these are a few that highlight the extra-research support needed and the advocacy role that NSF can play. These include:

- Create a mechanism to support long-term collaborations, especially across disciplinary boundaries
- Continue funding for RAPIDs that produced results, and encourage similar RAPID-type programs at other agencies
- Speed up proposal review process
- Foster improved science communication by organizing annual conferences that bring together media/press, professional communicators, influencers who can provide scientists with more effective communication strategies
- Provide funding for 2-year colleges
- Continue online research experiences for students (e.g., college does not have large research program, cannot travel, physical challenges)
- Educate regulatory agencies on potential impact of novel technologies
- Provide more opportunities and funding to allow for high school and collegiate real-world research

Several areas that could benefit from increased discussion through workshops or roundtables include:

- Electronic health records
- Politicization of medical issues
- Integration of social-behavioral-economic sciences into what we call the “hard” sciences
- Limitations of science to deal with things that are developing in real time
- Identification of which part of pandemic science has what level of confidence, and what are the knowns, what are the unknowns, what are the unknown unknowns
- Coordination with government agencies to implement pandemic response plans
- Issues surrounding rapid vaccine manufacture and distribution
- Consideration of data collection lessons learned in the creation of a global data infrastructure to face the next pandemic

Appendix A - Research Roadmap Recommendations

- Recommendation 1: Organize and/or fund initiatives that bring together policy makers, public health officials, and researchers from diverse disciplines with the primary objectives of building rapport and opening dialogue to build collaborative relationships.
- Recommendation 2: Develop coordination between US research funding agencies to ensure we have the capability to rapidly respond during a pandemic crisis.
- Recommendation 3: Advocate to increase the frequency and “typicality” of multidisciplinary research/collaboration at universities/research centers, funding agencies, and publication venues.
- Recommendation 4: Encourage development of an advanced computing cyberinfrastructure that is user-focused and ubiquitous, removing barriers to use and ensuring availability of dedicated resources in response efforts.
- Recommendation 5: Create a global data and computing infrastructure to archive cross-institution/ agency multimodal data, and analytic tools that were critical in mitigating the effects of the COVID-19 pandemic, allowing for rapid mobilization of these resources in times of crisis.
- Recommendation 5.1: Build a robust network of open synthetic data sources as an alternative to enhancing data sharing and access capabilities.
- Recommendation 6: Promote research that enhances and expands modeling capabilities, and support methods to communicate accurate interpretation to end users.
- Recommendation 7: Advocate for the establishment of a global biosurveillance infrastructure in conjunction with relevant government agencies that supports an international sample collection mechanism with long-term funding and unrestricted data sharing.
- Recommendation 8: Promote responsible data curation and use through supporting research that addresses bias and fairness issues, ultimately creating standards for data quality and facilitating equitable resource access.
- Recommendation 9: Create cross-NSF directorate research initiatives focused on identifying and mitigating the spectrum of societal disparities and impacts exposed and aggravated as a result of COVID-19.

Appendix B - NSF PREPARE Events

The PREPARE team collected this information from a number of sources, including three large virtual events (RP1, RP2, RP3) that addressed the full breadth of pandemic response research and five focused topical workshops. We met with our Steering Committee[17] as a group on five occasions, and had numerous independent conversations with various members. By hosting a podcast, *Science Before the Storm*[18], we were able to engage with several researchers on a more personal level. Hundreds of hours of presentations, discussions, and interviews contributed to the Research Roadmap; we have listed the events, number of participants, and links to the original material in this appendix.

Date	Event	Participants	Link
15-16 Dec 2020	RP1 (RAPID PIs)	149	RP1: Kick-off Workshop - YouTube
12-13 May 2021	Data and Computing Resources Workshop	337	Data and Computing Resources Workshop - YouTube
24-25 Jun 2021	Social, Behavioral, Economic, and Governance Aspects of Pandemics Workshop	253	SBEG: Social, Behavioral, Economic, and Governance Issues Workshop - YouTube
8-9 Dec 2021	RP2 (RAPID PIs)	222	RP2: 2nd Annual Rapid PI Meeting - YouTube
30-31 Mar 2022	HPC: Scalable Computing Workshop	159	HPC: Scalable Computing Workshop - YouTube
7-8 Jun 2022	Vaccine Preventable Diseases Workshop	335	Vaccine Preventable Diseases Workshop - YouTube
25-26 Jan 2023	LEVERS: Lessons & Experiences on Viable Epidemic Response Strategies	433	LEVERS Workshop - YouTube
13 Jul 2023	RP3 (Research for Pandemic Preparedness)	116	NSF PREPARE Research Roadmap
TOTAL		**2004**	

[17] https://prepare-vo.org/research-roadmap-next-pandemic
[18] https://open.spotify.com/show/49HJAQ6UxUIZM7shzbpA7V?si=f18d322811854780

Appendix C - Content Contributors

Many thanks to everyone who contributed to and participated in the workshops and podcasts hosted by the PREPARE team. This report was prepared with the input of over 2000 scientists, policymakers, and other stakeholders, and represents the synthesis of our discussions.

The people listed below served on program committees, moderated discussions, presented their work in talks and posters, actively participated in breakout sessions, and shared their insights through podcast interviews.

Name	Institution
H.B. Acharya	Rochester Institute of Technology
Mayank Agrawal	Princeton University
Dolores Albarracin	UIUC
Viggo Anderasen	Roskilde University
Sir Roy Anderson	Imperial College London
Taylor Anderson	George Mason University
Darryl Angel	Yale University
Jason Asher	CDC CFA
Hafiz Asif	Rutgers University
Pavel Atanasov	Pytho LLC
S. Farokh Atashzar	New York University
Chris Barrett	Biocomplexity Institute, UVA
Golda Barrow	Biocomplexity Institute, UVA
Amit Barui	Purdue University
Mitra Basu	NSF
Saikat Basu	South Dakota State University
Rupali Batta	Harvard University
Elizabeth Belding	UC Santa Barbara
Kenneth Bellian	Jensen Partners
Archna Bhatia	IHMC
Abhinav Bhushan	Illinois Institute of Technology
Matt Biggerstaff	CDC
Phil Blood	Pittsburgh Supercomputing Center
Glen Borchert	University of South Alabama
Sarah Bowman	Hauptman-Woodward Medical Research Institute

Thomas Breideband	University of California, Irvine
Shawn Brown	Pittsburgh Supercomputing Center
John Brownstein	Harvard University
Don Burke	University of Pittsburgh
Han Cai	MIT
Anton Camacho	MSF
Carolina Canovaro	WHO
Guohua Cao	Virginia Tech
Kathleen Carley	Carnegie Mellon University
Thomas Carton	Louisiana Public Health Institute
Kaushik Chakravarty	VeriSIM Life
Kartik Chandran	Columbia University
Xiang Cheng	Virginia Tech
Michael Chertkov	University of Arizona
Wai-Yim Ching	University of Missouri-KC
Haoran Chu	University of Florida
Monique Chyba	University of Hawaii
David Clark	MIT
Sarah Cobey	University of Chicago
Rita Colwell	University of Maryland
Angelique Corthals	CUNY John Jay College
Ingemar Cox	University of College London
Justin Crow	VDH
Jiaming Cui	Georgia Institute of Technology
Gautam Dasarathy	Arizona State University
Gautam Dasarthy	Arizona State University
Munmun DeChoudhury	Georgia Tech
Ashutosh Dhekne	Georgia Tech
Sai Manoj Pudukotai Dinakarrao	George Mason University
Kim Yi Dionne	UC Riverside
Stacey Doan	Claremont McKenna College
Jeffrey Donowitz	Virginia Commonwealth University
Alex Dow	Facebook
Dominique Duncan	University of Southern California

Alon Efrat	University of Arizona
Rosalind Eggo	LSHTM
Joshua Epstein	NYU
Mark Escott	Austin EMS
Stephen Eubank	Biocomplexity Institute, UVA
Nina Fefferman	University of Tennessee-Knoxville
Stuart Feldman	Schmidt Futures
Matt Ferrari	Penn State
Mark Fichtel	University of Oklahoma
Ken Fleischmann	University of Texas-Austin
Tina Fletcher	Florida International University
Sebastian Funk	LSHTM
Borko Furht	Florida Atlantic University
Evgeniy Gabrilovich	Google Health
Lazaros Gallos	Rutgers University
Aryya Gangopadhyay	UMBC
Niu Gao	Public Policy Institute of California
Song Gao	University of Wisconsin-Madison
Rachael Garner	University of Southern California
Nora Garza	Laredo College
Erwin Gianchandani	NSF
Julia Gog	Cambridge University
Jennifer Golbeck	University of Maryland
Raju Gottumukkala	University of Louisiana, Lafayette
Pelagia Gouma	The Ohio State University
Andrey Grigoriev	Rutgers University
Wendy Guan	Harvard University
Katie Hail-Jares	Griffith Criminology Institute
Sharika Hegde	Northwestern University
John Heidemann	Information Sciences Institute, University of Southern California
Andy Hock	Cerebras
Deidre Hollingsworth	Oxford University
Suellen Hopfer	University of California, Irvine

Eric Horvitz	Microsoft
Florence Hudson	Columbia University
Nathaniel Hupert	Weill Cornell Medicine
Amit Huppert	Tel Aviv University
Mac Hyman	Tulane
Constance Jeffery	University of Illinois, Chicago
Michael Johansson	CDC
Kyle Johnsen	University of Georgia
Neil Johnson	George Washington University
Kirk Jordan	IBM
Regina Joseph	Pytho LLC
James Joshi	NSF
Wendy Ju	Cornell Tech
Murat Kantacioglu	UT Dallas
Murat Kantarcioglu	University of Texas, Dallas
Sanjiv Kapoor	Illinois Institute of Technology
Sreenitha Kasarapu	George Mason University
Kasra Khademorezaian	University of Oklahoma
Michael Klaczko	University of Rochester
Eili Klein	Johns Hopkins University
Alice Koniges	University of Hawaii
Archanaa Krishnan	Virginia Tech
Vasileios Lampos	University of College London
Heidi Larson	LSHTM
David Lazer	Northeastern University
Christian Lebiere	Carnegie Mellon University
Peter Leonardi	OmniCyte
Jure Leskovec	Stanford University
Simon Levin	Princeton University
Bryan Lewis	Biocomplexity Institute, UVA
Chen Li	University of California, Irvine
Lily Li	Biocomplexity Institute, UVA
Ninghui Li	Purdue University
Ji Lin	MIT

Yu-Ru Lin	University of Pittsburgh
Yang Liu	LSHTM
Chenghua Long	Columbia University
Cristina Lopes	University of California, Irvine
Ben Lopman	Emory University
Yung-Hsiang Lu	Purdue University
Maimuna Majumder	Harvard Medical School
Ilan Mandel	Cornell Tech
Achla Marathe	Biocomplexity Institute, UVA
Madhav Marathe	Biocomplexity Institute, UVA
Gloria Mark	University of California, Irvine
Margaret Martonosi	NSF
Michael Matthews	University of Oklahoma
Kerrigan McCarthy	NICD
Mimi McClure	NSF
Deep Medhi	NSF
Sharad Mehrotra	University of California, Irvine
Antonio Merolli	Rutgers University
Jessica Metcalf	Princeton University
Lauren Meyers	University of Texas-Austin
Maria Jose Miguez	Florida International University
Danil Mikhailov	data.org
Bud Mishra	NYU
Dina Mistry	Twitter
Finn Mokrzycki	UVA
James Moody	Duke University
Anthony Morciglio	Georgia State University
Jonathan Mosser	IHME/UW
Thumbi Mwangi	University of Nairobi
Giri Narasimhan	Florida International University
Krishna Narayanan	Texas A&M University
Satish Narayanasamy	University of Michigan
Kartik Nayak	Duke University
Wilfred Ndifon	African Institute for Mathematical Sciences

Kien Nguyen	University of Southern California
Wendy Nilsen	NSF
Lucila Ohno-Machado	UCSD
Timothy Oladunni	University of the District of Columbia
Marie Jo Ondrechen	Northeastern University
Eyal Oren	San Diego State University
Mark Orr	Biocomplexity Institute, UVA
Philip Pare	Purdue University
Laxmi Parida	IBM
Ioannis Paschalidis	Boston University
Sameer Patil	Indiana University Bloomington
Peter Pirolli	IHMC
H. Vincent Poor	Princeton University
Anne Marie Porter	American Institute of Physics
Jim Powell	University of Alaska Southeast
B. Aditya Prakash	Georgia Tech
William H. Press	University of Texas-Austin
Calton Pu	Georgia Institute of Technology
Juliet Pulliam	SACEMA
Celia Quinn	NYC DoHMH
Ali Rahnavard	George Washington University
Arvind Ramanathan	Argonne National Laboratory
Bhargavi Rao	LSHTM/MSF
Praveen Rao	University of Missouri-Columbia
Ramesh Raskar	MIT
Andy Rasmussen	Chicago Public Schools
S.S. Ravi	Biocomplexity Institute, UVA
Indrajit Ray	NSF
Indrakshi Ray	Colorado State University
Erin Raymond	Biocomplexity Institute, UVA
Beth Redbird	Northwestern University
Nick Reich	University of Massachusetts-Amherst
Philip Resnik	University of Maryland
Nigel Reuel	Iowa State University

Amira Roess	George Mason University
Dan Romer	Penn State
Arif Sadri	Florida International University
Marcel Salathé	EPFL
Daniel Salmon	Johns Hopkins University
Sam Scarpino	Northeastern University
John Seberger	Michigan State University
Cyrus Shahabi	University of Southern California
Bahar Shahverdi	George Mason University
Jeffrey Shaman	Columbia University
Shantanu Sharma	University of California, Irvine
Sharad Sharma	Bowie State University
Katriona Shea	Penn State
Chi-Ren Shyu	University of Missouri, Columbia
Lone Simonsen	Roskilde University
Gurdip Singh	NSF
Vivek Singh	Rutgers University
Paul Slovic	University of Oregon
Davida Smyth	Texas A&M University-San Antonio
Gabriela Solis-Cavazos	Laredo College
Xiao Song	University of Southern California
Anna Squicciarini	Penn State
Ashok Srinivasan	University of West Florida
Ajitesh Srivastava	University of Southern California
Raphael Stern	University of Minnesota
Rick Stevens	Argonne National Laboratory
Fred Streitz	CDC/LLNL
Deborah Strumsky	University of Maine-Center for Business and Economic Research
Heshan Sun	University of Oklahoma
Rajesh Sundaresan	Indian Institute of Science
Abhijit Suprem	Georgia Tech
Raji Tajudeen	Africa CDC
Milind Tambe	Harvard University

Sylvia Thomas	University of South Florida
John Towns	UIUC
Shaun Truelove	Johns Hopkins University
Salil Vadhan	Harvard University
Jaideep Vaidya	Rutgers University
Aram Vajdi	Kansas State University
Moshe Vardi	Rice University
Srini Venkatramanan	Biocomplexity Institute, UVA
Alessandro Vespignani	Northeastern University
Venkat Vishwanath	Argonne National Laboratory
Anil Vullikanti	Biocomplexity Institute, UVA
Fei Wang	Weill Cornell Medicine
Wei Wang	UCLA
Krista Wigginton	University of Michigan
Dmitri Williams	University of Southern California
Matthew Wolfgram	University of Wisconsin-Madison
Joseph Wu	Hong Kong University
Xiaochen Xian	University of Florida
Li Xiong	Emory University
Lanyu Xu	Wayne State University
Chaowei Yang	George Mason University
Yaling Yang	Virginia Tech
Cao Yifeng	Georgia Institute of Technology
Hongwei Yu	University of Wisconsin-Madison
Jin Yu	University of California, Irvine
Elena Zheleva	University of Illinois, Chicago
Andreas Züfle	Emory University