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 our nation's response to the next pandemic. Since we started this project in October 2020, we have hosted seven virtual workshops featuring 72 subject-matter experts as speakers, panelists, and committee members. Collectively, these sessions were attended by over 2200 researchers and viewed on YouTube more than 2600 times. Please see https://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 Corporate-Government-Academic Partnerships and Multidisciplinary Collaboration; Advanced Computing and AI; Data and Computing Infrastructure; Tools and Methods; and Societal Impacts.

The recommendations offered in this Research Roadmap can act as an effective tool for NSF leadership, and significantly benefit international research initiatives focused on pandemic resilience and preparedness.

OVERVIEW

During the summer of 2020 as COVID-19 ravaged the world, 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) 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 variety 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 dealing with future pandemics. The Research Roadmap includes the identification of key research topics, risks, and gaps in the current R&D landscape that will 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.

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 subject-matter experts as speakers, panelists, and organizers of these workshops. Our total virtual attendance number among all sessions is over 2200 people, with an additional 3612 views on YouTube. We continue to engage with several participants after the workshops, keeping the dialogue alive within the community. We also continue to interact with community members through our *Science Before the Storm* podcast, where we have enjoyed conversations with several scientists in the first two

seasons and will release our third season in Spring 2023. And finally, we have looked at a variety of reports and publications to further inform our recommendations[1].

The research community is still very invested in the topic of pandemic preparedness, and a high attendance at our most recent workshop highlights the interest level. Our LEVERS (*Lessons and Experiences on Viable Epidemic Response Strategies*) workshop in January 2023 had 401 registrations, and has been viewed on YouTube (@nsf_prepare) 800+ times. 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 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 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 (for more information 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 Corporate-Government-Academic Partnerships and Multidisciplinary Collaboration; Advanced Computing and AI; Data and Computing Infrastructure; Tools and Methods; and Societal Impacts. We have also included holistic recommendations for NSF and our ideas for future workshops PREPARE can organize to collect additional input for the Research Roadmap.

---

[1] 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

I. Importance of Corporate-Government-Academic Partnerships and Multidisciplinary Collaboration

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 and 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 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

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. Additionally, trust must be built between public policy makers and minority communities; citizens must trust that there is no ulterior motive for interventions. Communicating uncertainty is critical to informed decision making.

> **Recommendation:**
> **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.**

With interactions of this nature, 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.

National Agencies, Corporations, and Foundations

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 leads 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.

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

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

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

Mechanisms to collect data during, but also before and after the pandemics and mechanisms to maintain them. RAPID grants served as an excellent mechanism to collect data and develop code during the pandemic. But now that the pandemic has ended this effort needs to continue. Further data collected by these programs needs to be made easily accessible and maintained; lack of continued funding often leads to the inaccessibility of these data sets over time. Journals, NSF centers and repositories have played a constructive role in this direction thus far.

Multidisciplinary collaboration

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.

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

- Encourage and fund long 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 wide variety of opportunities for collaboration, varying from videographers to geneticists. 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. Overcoming the challenge of finding a common multidisciplinary language will enhance the complex collaborations in pandemics science today that require different disciplines to effectively communicate and work together on transdisciplinary research directions.

II. Advanced Computing and AI

While advanced computing initiatives are critical within the 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:**
> **Encourage development of an advanced computing and AI ecosystem 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 those whose work might have benefited from using HPC found the barriers to adoption to be very high. Non-traditional HPC users, like AI and machine learning epidemiology, require service-oriented architectures that allow one to connect to different resources in novel ways, and modifying workflows to fit the parameters of the HPC center was an onerous burden. Users do not always know in advance the exact resources they will need to complete their work, and they should be provided with resources to allow them that flexibility. Some ways to improve HPC adoption by the research community include:

- Creating the infrastructure before any new outbreaks, removing any perceived impediments to access (e.g., any extra paperwork or proposals), providing a dedicated customer support system, providing a dedicated virtual container for development, debugging and runtime, and perhaps providing incentives to participate
- Building software platforms that are focused on the research work rather than the low-level details of how to run codes on HPC resources
- Addressing 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

- Educating emerging communities about advanced computing resources, what is available, and how it can apply to their research
- Resolving the challenges of moving a workflow to another system

III. Data and Computing Infrastructure

While many researchers acknowledge the benefits of access to an unprecedented amount of data during the COVID-19 pandemic, the sheer variety of originating sources and multimodal types revealed challenging practical use issues. It's 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, the following recommendations could prove helpful in overcoming these challenges.

> **Recommendation:**
> **Create a global data and computing infrastructure to archive cross-institution/ agency multimodal data, analytic tools, and hardware solutions that were critical in mitigating the effects of the COVID-19 pandemic, allowing for rapid mobilization of these resources in times of crisis.**

Within this infrastructure, researchers could: develop hardware solutions for biosurveillance; build privacy-preserving and secure data collection, sharing, and analysis tools; and create pilot testbeds where researchers can evaluate privacy-enhancing technologies. The scientific community must better understand which data was missing from the COVID-19 analysis that could have had the biggest impact on mitigation. 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; mobility; behavioral; county-level NPI, vaccine allocation, confirmed case counts, and hospitalizations; and a nationwide longitudinal cohort study.

> **Recommendation:**
> **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.**

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 and so the development of privacy technology with appropriate governance and procedural control must be encouraged. 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 1 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 (Findability, Accessibility, Interoperability, and Reusability) and accessible, and funding mechanisms for facilitating, finding, and aggregating related data to open platforms for analysis.

Recommendation:
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 and preserving individual privacy. One potential path through this roadblock is the creation of multiple substantial and diverse synthetic datasets. 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. 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[2].

[2] Harrison, G., Chen, J., Mortveit, H., Hoops, S., Porebski, P., Xie, D., Wilson, M., Bhattacharya, P., Vullikanti, A., Xiong, L., Marathe, M. (2023). 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

IV. Tools and Methods

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 modeling and biosurveillance, which were generally identified in our workshops as highly impactful in pandemic response.

Modeling

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.

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

One example of an ensemble modeling method that proved very useful to policy makers was the COVID-19 ForecastHub. 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[3].

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

- Creation of methodological advances to bridge AI/ML and develop multi-scale multi-theoory models that take behavioral/causal, cognitive, political and economic, immunological and virological issues into account
- Development of 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
- Increased use of modeling to inform biosurveillance
- Creation of benchmarks to compare models
    - Predictive or retrospective validity of models, although useful, is not adequate
    - Model utility should also be evaluated in terms of its ability to support effective decision making, situation assessment, and counterfactual analysis
- Documentation of model reliability to build trust

Pervasive Sensing and Biosurveillance

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

[3] 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.

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.

Recommendation:
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.

We understand that this recommendation may fall outside the scope of NSF's mission, but it is a critical need. To that end, NSF can support research in these areas:

- Building a pathogen testing infrastructure that can detect the emergence of novel pathogens, similar to early warning systems for detecting earthquakes
- Creating a more robust surveillance infrastructure, new surveillance apps, and effective tools for finding threats
- Developing technology and encouraging the implementation of policy for efficient ways to do contact tracing and quarantine infrastructure
- Creating 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

V. 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:**
> **Create cross-NSF directorate research initiatives focused on identifying and mitigating the spectrum of societal disparities and impacts exposed as a result of COVID-19.**

Researchers who attended 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 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 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

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

Future Workshops

- (July 2023) RP3 - Research for Preparedness - we plan to unveil the penultimate draft of the Research Roadmap to the community and seek their input for the final version.
- (Fall 2023) By bringing together a diverse group of researchers, we plan to explore the opportunities and limitations surrounding electronic health records and their role in pandemic response. We will discuss: data accessibility, aggregation, standardization, and linkage; ML and AI usage in this space; and applications to health services to address future emergency response needs and ongoing concerns like long COVID.

Potential Future Workshops

- Politicization of medical issues
- Integration of social-behavioral-economics 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
- Scientific communication