

Harvard Data Science Review • Issue 3.1, Winter 2021

We Need a (Responsible!) Data Science Rapid Response Network

Eric Kolaczyk, Meredith M. Lee, Jing Liu, Micaela S. Parker

Published on: Jan 29, 2021

DOI: 10.1162/99608f92.2794e78d

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

During the COVID-19 pandemic, for better or worse, we are learning much about ourselves as a society. Sabina Leonelli (“Data Science in Times of Pan(dem)ic,” this issue) provides a broad, cogent, and thought-provoking reflection on much of what we have learned to date from a data science perspective. Her chosen device for doing so—through imaginaries of data use—is particularly effective (and refreshingly unfamiliar to most data scientists) in balancing a sense of both what has been and what might be. The imaginaries include a welcome framing of the pitfalls of two broad areas of data science contributions during the pandemic: population surveillance and predictive modeling.

Rather than similarly focusing on “the ways in which the data science contributions to the pandemic response are imagined and projected into the future,” here we center and emphasize the question of appropriate mechanisms for facilitating, or ideally even optimizing, such contributions. We advocate for mechanisms that will yield a coordinated and responsible *rapid* response, not a *rushed* response, from the data science community in times of crises, steeped in an evolving understanding of societal context. Specifically, we make the following points: 1) A data science rapid response network with a generalizable framework would help address not only gaps illuminated by the current pandemic, but also anticipated challenges in many types of future crises; 2) A data science rapid response network should leverage existing consortia, centers, and networks—strengthening capacity locally, regionally, and more broadly; and 3) A data science rapid response network can bring together talent, support scientific integrity, and help translate scientific discoveries to action in unique ways, facilitating relevant and timely coordination.

Reflecting on Lessons Learned

The enormous collective data science efforts across the country, and the world more generally, have supported and informed much of society’s management and control of COVID-19. The data science community engaged quickly, with contributions that ultimately led to new insights into nearly every facet of the pandemic. Examples range from online tools enabling 3D exploration of genomic variants of coronavirus (Portelli et al., 2020), to analyses that uncover estimation biases in fatality rates (Angelopoulos et al., 2020) and the impact of lockdown strategies in different countries mitigating these rates (Pachetti et al., 2020), to combating misinformation about COVID-19 propagated in news outlets and on social media (reviewed in Starbird et al., 2020). The COVID Information Commons (CIC), established with funding through the United States National Science Foundation (NSF) RAPID response research program to enable collaboration across hundreds of researchers working on the pandemic, provides nearly 1,000 additional examples of such contributions funded by NSF alone.

Yet, scores of groups eager to contribute to data-driven efforts have not been effectively connected with each other. This was particularly evident at the outset of the pandemic, when *ad hoc* teams of colleagues and local networks seemed to spring up across the country and, indeed, the world. More

ambitious groups pulled together global-scale websites, crowdsourced lists of resources, and hosted volunteer matching opportunities. But these too struggled to keep up with the pulse of the pandemic and were not themselves well connected.¹ The end result is that knowledge and insight gained from these collective efforts often lagged weeks and months behind community needs during the global spread of the virus. Pockets of innovation continue to sprout up across the country and the world; whether focused on data, models, vaccines, analysis, or discoveries, coordination and cohesion remain elusive.

In other words, while the data science community responded as rapidly to COVID-19 as it was able—offering dashboards, preprints, and an explosion of research projects at an unprecedented pace—the end result nonetheless was not a *rapid response*. Notably absent was a shared infrastructure for data science that can interface with existing disaster response efforts and local stakeholders, to ensure that data-enabled tools and approaches are well-matched to the real needs of frontline workers, first responders, and the underserved and underresourced communities often most impacted by disasters.

We Can Do Better

To prepare for the next inevitable crisis—whether regional, national, or global in scope, and whether focused on health, natural disasters, or humanitarian issues—the data science community needs a trusted, dependable, flexible, and cohesive data-enabled rapid response network. The COVID-19 pandemic has demonstrated that we require more from data science than relatively isolated contributions and impromptu consortia. Our collective ability to use data science optimally in addressing societal needs increasingly hinges on a convergence of agility and coordinated capacity, with academia, government, industry, and community organizations each bringing complementary resources to bear. What is needed is the practiced usage of shared infrastructure, with consistent architecture, preparedness training, and interoperable teams—working in concert with the future of emergency response networks, data providers, and data users through clear and efficient communication channels.

An effective data science rapid response network is unlikely to emerge fully and well-formed without concerted effort. Scientifically, the undertaking is inherently interdisciplinary. It will require key contributions from core computational disciplines that are crucial to addressing challenges around hardware and software infrastructure, cloud computing, data and model sharing, engineering, analytics, and artificial intelligence (AI). Subject to the type of disaster, equally important will be the engagement and contributions of disciplines spanning the natural and social sciences, public health, urban planning, and more. Fundamental insights from organizational management, education, communication, ethics, policy, and law will be critical to incorporate as well.

Organizationally, the tent of data science is positioned at the intersection of a spectrum of different communities of scientific expertise, across academic, government, industry, and nonprofit sectors—an intersection that is still in the early stages of developing a sense of itself as a community. At the same time, much of the culture and spirit of data science has been one of innovation, exploration, disruption, and change. This spirit can be conducive toward progress in a host of areas that have already profoundly transformed society, but is lacking precisely in many of the characteristics that define robust emergency response. Whether implicitly or explicitly acknowledged, the collective question before us is how to learn from the pandemic’s continuing damage, and whether the data science community can and will do something different. While the community has clearly demonstrated interest, sustainable organizational leadership and commitment are needed to fully realize the potential that data science can offer these large societal challenges.

Facilitating Next Steps

Fortunately, in the United States we already have many key elements of what could form the foundation for a rapid response network. Nearly every major research university now has, or will soon have, a data science ‘institute’ or other academic unit (Parker et al., in press). Similarly, the NSF-funded regional Big Data Innovation Hubs already facilitate collaboration at the confluence of research, education, and practice, developing partnerships across sectors for societal needs. And groups like the Academic Data Science Alliance (as well as a constellation of related professional societies) are working to help ensure the quality and integrity of data science research and education, in addition to communicating with the larger community of stakeholders and guiding policy. The COVID-19 crisis stress-tested the infrastructure that we have, and found it wanting. But it also showed us the potential for what could be. This is the time for us to solidify our foundation and build a well-oiled infrastructure that can withstand the test of the next crisis.

Key to regional and national efforts is that this infrastructure be able to coordinate the development of trusted multisector partnerships and deliverables, engage with existing disaster response and management organizations, and leverage insight from data science institutes and communities. Cooperation with data and technology partners in the private sector will heighten the network impact and scale; continued collaboration that includes social services and local community organizations will be essential to address equity and social justice issues that are amplified during disasters.

We are far from the only ones to call for a greater degree of coordination and infrastructure, both across the data science community and beyond, in response to lessons learned during the current pandemic. Examples range from frameworks for the hazards and disasters field (Peek et al., 2020) to overviews in the field of ethics (McGuire et al., 2020), and broad reimaginings of a new data-enabled future (Lee et al., 2020), as well as related calls for notions like ‘Science Readiness Reserves’ (Loeb & Gil, 2020). Our own efforts in recent months have shifted toward trying to facilitate the type of change

we envision above. These were kicked off with an October 14, 2020, joint session of the [ADSA Annual Meeting](#) and the [Data Science Leadership Summit](#), intended to act as a launching point for partnership commitments and future coordination to establish the foundation and strategic path toward a data science rapid response to crises.²

Participants of the joint session envisioned a data science rapid response network under a generalizable framework that can adapt to uncertain environments and be effective for any type of crisis—not just public health ones—with processes, technologies, and relationships that are solidified, verified, and stress-tested well beforehand. This network will need to be well connected with other rapid response and disaster relief organizations, alongside a broad alliance of stakeholders. It must bolster and maintain the relationships that we can rely on when disasters hit, with open channels between academia, industry, government, and the whole community. The network should aim to incorporate the agility and context of local networks serving the unique needs of local stakeholders, yet bridge to the larger scope of nationwide coordination and unified larger goals. The network should play an essential role in three areas.

1) *People*. The network must build and sustain the ‘connectivity’ needed for people to come together effectively during a crisis. It must connect data scientists with policymakers, emergency management leaders, and community stakeholders *before* a crisis to understand where goals, capacities, and interests align. The network could facilitate ‘surge staffing’ through advocacy for culture change at universities and other sectors, to allow a ‘reserve’ of scientists to walk away from their regular duties (grants, publications, tenure clocks) and devote themselves to rapid response research activities. It could also offer disaster response training for data scientists and students, and data science literacy for policymakers, disaster response workers, and community stakeholders for effective collaboration with data scientists.

2) *Science*. The network must facilitate the development of trusted processes for data collection, management, use, and protection, and put in place collaborative agreements, data-sharing agreements, and similar arrangements well in advance of a crisis. The network should facilitate adoption and support development of unified data collection standards and definitions, and methods to resolve or address disagreements between data at different levels (e.g., local vs. national levels). It should help ensure that the rapid response science is good science through coordinated data curation, well developed workflows, and mechanisms for rapid peer review. The network should also play a critical role to ensure that the rapid response science has integrity, by building data science ethics expertise into the network, and highlighting data sovereignty and privacy issues in the context of crisis response goals.

3) *Translation*. The network must support translating the science into informed dialogue, guidance, policies, and action with transparency and accountability, through two-way communication that

recognizes history and context among stakeholders, feeding back into the data science. The network could also play a role in building trust among scientists, policymakers, the disaster response community, and the broader public—a trust that in turn rests on trust in the integrity and ethics of the underlying science. It must help ensure equity and inclusion for vulnerable and underserved populations, especially when research is translated into policy.

Our plan is to build upon the energy of the October event to help convene stakeholders and resources—understanding existing infrastructure, gaps, and pain points—to lay further foundation and a path forward. We envision that a 2-year effort could, with a representative cross-section of academic, private, and public sectors, (i) convene stakeholders and build trust around shared values and goals—noting relevant prior work, insights, and resources that can be extended and adapted; (ii) establish protocols for more efficient and effective data-informed crisis communication networks; (iii) develop replicable materials for preparedness training units; and (iv) stand up situation-adaptive infrastructure and processes for two to three pilot use cases. We envision that with a well-supported 5-year effort, proof of concept of adaptation, adoption, and scaling to regional levels is possible.

In all likelihood, we will not have the luxury of 5 years without a crisis to ‘get our house in order’ as a data science community. The reality is that there is probably a crisis somewhere in the world every day. But as we imagine our shared future in a broader societal backdrop, what we *can* do is to reflect deeply on the levers we have and the learnings we need to continue supporting—and build now what we know we will need later.

Acknowledgments

Portions of this discussion are adopted from Kolaczyk, E. (2020). POV: COVID-19 Shows Us We Need Rapid Response Data Science Teams. BU Today. June 10, 2020 <http://www.bu.edu/articles/2020/pov-covid-19-rapid-response-data-science-teams>.

Disclosure Statement

Portions of this work, including that deriving from the ADSA Data Science Leadership Summit and Annual Meeting, were supported by the Gordon and Betty Moore Foundation (grant #8432 to MSP), the Alfred P. Sloan Foundation (grant #G-2019-11447 to MSP), and the National Science Foundation (grant #2034493 to MSP; grant #1916573 to MML). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Angelopoulos, A. N., Pathak, R., Varma, R., & Jordan, M. I. (2020). On identifying and mitigating bias in the estimation of the COVID-19 case fatality rate. *Harvard Data Science Review (Special Issue 1)*.

<https://doi.org/10.1162/99608f92.f01ee285>

Lee, M. M., Johnson, A. D., Yelick, K. A., & Chayes, J. T. (2020). The road for recovery: Aligning COVID-19 efforts and building a more resilient future. *IEEE Data Engineering Bulletin*, 43(2), 133–140.

<http://sites.computer.org/debull/A20june/p133.pdf>

Loeb, A., & Gil, D. (2020, April 30). Let's create an elite scientific body to advise on global catastrophes [Op Ed]. *Scientific American*. <https://blogs.scientificamerican.com/observations/lets-create-an-elite-scientific-body-to-advise-on-global-catastrophes/>

McGuire, A. L., Aulisio, M. P., Davis, F. D., Erwin, C., Harter, T. D., Jagsi, R., Klitzman, R., Macauley, R., Racine, E., Wolf, S.M., Wynia, M., Wolpe, P.R., and The COVID-19 Task Force of the Association of Bioethics Program Directors (ABPD) (2020). Ethical challenges arising in the COVID-19 pandemic: An overview from the Association of Bioethics Program Directors (ABPD) task force. *The American Journal of Bioethics*, 20(7), 15-27. <https://www.tandfonline.com/doi/full/10.1080/15265161.2020.1764138>

Pachetti, M., Marini, B., Giudici, F., Benedetti, F., Angeletti, S., Ciccozzi, M., Masciovecchio, C., Ippodrino, R., & Zella, D. (2020). Impact of lockdown on Covid-19 case fatality rate and viral mutations spread in 7 countries in Europe and North America. *Journal of Translational Medicine*, 18, Article 338.

<https://doi.org/10.1186/s12967-020-02501-x>

Parker, M. S., Burgess, A. E., & Bourne, P. E. (in press). Ten simple rules for starting (and sustaining) an academic data science initiative. *PLOS Computational Biology*. OSF Preprint.

<https://doi.org/10.31219/osf.io/wu4fv>

Peek, L., Tobin, J., Adams, R., Wu, H., & Mathews, M. (2020). A framework for convergence research in the hazards and disaster field: The natural hazards engineering research infrastructure CONVERGE facility. *Frontiers in Built Environment*, 6, [110]. <https://doi.org/10.3389/fbuil.2020.00110>

Portelli, S., Olshansky, M., Rodrigues, C. H. M., D'Souza, E. N., Myung, Y., Silk, M., Alavi, A., Pires, D. E. V., & Ascher, D. B. (2020). Exploring the structural distribution of genetic variation in SARS-CoV-2 with the COVID-3D online resource. *Nature Genetics*, 52, 999–1001. <https://doi.org/10.1038/s41588-020-0693-3>

Starbird, K., Spiro, E.S., & Koltai, K. (2020, June 25). Misinformation, crisis, and public health—Reviewing the literature. *MediaWell*. <http://doi.org/10.35650/MD.2063.d.2020>

This discussion is © 2021 by the author(s). The editorial is licensed under a Creative Commons Attribution (CC BY 4.0) International license (<https://creativecommons.org/licenses/by/4.0/legalcode>), except where otherwise indicated with respect to particular material included in the article. The article should be attributed to the authors identified above.

Footnotes

1. A living list of such resources is available at the [Academic Data Science Alliance's COVID-19 Resources webpage](https://academicdatascience.org/covid) (<https://academicdatascience.org/covid>). ↵
2. The video of the panel discussion, as well as more information on our efforts, can be found at: <https://academicdatascience.org/programs-and-communities/rapidresponse-datascience>. ↵