NSF Convergence Accelerator
Privacy and Pandemics: Responsible Use of Data During Times of Crisis

Executive Summary

On October 27 and 28, 2020, the Future of Privacy Forum (FPF) convened an international group of computer science, privacy law, social science, and health information experts in a workshop titled, “Privacy and Pandemics: Responsible Use of Data During Times of Crisis,” to examine benefits, risks and strategies for the collection and protection of data in support of public health initiatives specific to COVID-19 and beyond. With support of the National Science Foundation, the workshop was organized and designed to identify research priorities to improve data governance systems and structures for future pandemics and help set direction for NSF’s 2021 Convergence Accelerator program.

Organizing Approach and Multi-Stakeholder Participation

FPF issued a worldwide “call for positions” to bring the best thinking of technologists, scientists, policymakers, data experts, industry leaders, and others together to consider how data and technology have each played a role in efforts to study, control the spread of, and track COVID-19 and the challenges and information gaps therein. Fifty-one submissions were selected for inclusion in the workshop and organized into four sessions across tech-policy-legal-ethical dimensions:

- Accessibility of Data to Track SARS CoV-2 Infections
- Use of Technology to Track, Trace & Notify to Control Spread of COVID-19
- Adapting Legal and Regulatory Responses to a Global Emergency
- Convergence of Technology, Policy and Responsible Data Use in a Global Crisis

Invited participants represented a range of disciplines and organizations from around the world, including scholars from Australia, the United States, India, Israel, and the European Union. Keynote presentations by Dr. Lauren Gardner, Creator of Johns Hopkins COVID-19 dashboard, and UC Berkeley Data Analytics Researcher Dr. Katherine Yelick spoke to variable issues of data access, data sharing, and data quality.

FPF used its virtual conference platform (Zoom) to feature presenters in moderated discussion with FPF experts before a large audience drawn from FPF’s academic-industry-government networks. (On both October 27 and 28, approximately 300 unique individuals attended at least one workshop session.)

Summary Points and Roadmap

Based on FPF’s analysis of leading research and expert positions, we recommend that NSF consider the following roadmap for research directions, practice improvements, and development of privacy-preserving products and services to inform responses to the COVID-19 crisis and in preparation for future pandemics and other crises:

- Support the refinement and application of existing privacy-preserving and privacy-enhancing technologies that can support public health goals while mitigating privacy risks. Promising approaches include: decentralized contact tracing, homomorphic encryption, and differential privacy;
- Support the development of emerging privacy-enhancing technologies that hold promise in the public health sphere. Emerging technologies include: synthetic data, controlled access environments, digital twins, and simulations;
- Support cross-disciplinary research into privacy-protective approaches to key emerging technologies, including Wireless Sensor Networks and data processing strategies for on-device and/or centralized analysis of personal health information;
- Explore mechanisms to balance the need for increased access to data that allows researchers to understand the differential impacts of crises on certain communities by not obscuring critical community characteristics with privacy-enhancing technologies;
- Convene cross-disciplinary experts to create and refine guidance for implementation of privacy protections suited to crisis situations;
- Identify top-priority updates to laws and regulations pertaining to public health;
- Explore mechanisms that promote data interoperability while promoting privacy;
- Promote the development of promising de-identification technologies and mitigation strategies to address re-identification risks;
- Promote practical, implementable ethical frameworks that go beyond the FAIR principles; and
- Identify practical lessons learned during the COVID-19 pandemic regarding publication ethics and norms for research in a time of crisis that can apply to future crises.

These many points were discussed in detail and at length. Other adjacent points and elaborations are reflected in the final draft report. (Note: All workshop participants were invited to comment on an early draft. The final draft report was reviewed by key stakeholders prior to submission.)

C-Accel Proposal

FPF has proposed Privacy and Pandemics: Responsible Use of Data During Times of Crisis as the focus of NSF’s latest Convergence Accelerator (“C-Accel”) program. Our C-Accel proposal is driven by urgent need to develop effective structures, protocols and processes for data sharing and governance while protecting individual privacy in a post-pandemic world.

In a three-year time frame, specific endeavors could include establishing convergence teams to develop best-practices guidance and data-based policy formulation for responsible data use in support of health initiatives; encouraging interoperability of the many data sources that inform data-driven healthcare decision-making; creating FAIR data repositories; facilitating multi-institutional sharing of related data science education and research expertise; and accelerating programs for ethical sharing of data across industry-academic collaboratives.

The proposed track will leverage and expand NSF's other COVID-related investments toward a goal of improved data sharing and governance with individual privacy rights and protections at the forefront. The Privacy and Pandemics workshop illustrated the value of shared forums like this to define and marry technology needs with good policy/policy frameworks. Both are needed to accomplish system-level changes and achieve long-term sustainability.

NSF Convergence Accelerator
Privacy and Pandemics: Responsible Use of Data During Times of Crisis
(C-Accel 2035358)

Jules Polonetsky, CEO, Future of Privacy Forum
Dr. Sara Jordan, Policy Counsel-- Artificial Intelligence, Future of Privacy Forum
Christy Harris, Director of Technology and Privacy Research, Future of Privacy Forum
John Verdi, Vice-President, Future of Privacy Forum

1.0 Background

In our 2019 paper "The Future of Privacy Technology," the Future of Privacy Forum (FPF) argued that "the technical, organizational, and logistical complexity of modern-day data governance and the need to protect and respect individual privacy rights constitute a 'grand challenge problem' that will require a grand challenge-level focus and investment to solve."[1] Subsequent to the emergence of the COVID-19 pandemic crisis, this position has crystalized and become more urgent. The pandemic has demonstrated that "data governance remains *the* modern business and societal problem."[2]

Based upon the insights gleaned from our Privacy and Pandemics conference[3] and related work by FPF, we lay out a case for doing research that addresses responsible uses of technology and data during crisis situations, a first-order priority of the National Science Foundation (NSF) and adjacent US federal funding bodies.[4]

Based on FPF's analysis of leading research and expert positions, we recommend that NSF consider the following roadmap to point the way forward in research directions, practice improvements, and development of privacy-preserving products and services to inform responses to the COVID-19 crisis and in preparation for future pandemics and other crises:

- Support the refinement and application of existing privacy-preserving and privacy-enhancing technologies that can support public health goals while mitigating privacy

[1] Future of Privacy Forum. (2019). "The Future of Privacy Technology." Available at: https://fpf.org/wp-content/uploads/2020/03/NSF_FPF-REPORT_C-Accel1939288_Public.pdf

[2] This material is based upon work supported by the National Science Foundation under Grant No. 2035358.

[3] "Privacy and Pandemics: Responsible Uses of Technology and Health Data During Times of Crisis—An International Tech and Data Conference" (Oct 27-28, 2020), https://fpf.org/2020-pandemics-conference/.

[4] "Corporate Data Sharing Workshop" (Mar. 26, 2020), https://fpf.org/2020/03/27/privacy-and-pandemics-a-thoughtful-discussion/; "A Closer Look at Location Data" (Mar. 25, 2020), https://fpf.org/2020/03/25/a-closer-look-at-location-data-privacy-and-pandemics/; "The Role of Mobile Apps" (Apr. 6, 2020), https://fpf.org/2020/04/14/privacy-pandemics-virtual-workshop-the-role-of-mobile-apps/; "Student Privacy During the COVID-19 Pandemic" (Mar. 20, 2020), https://ferpasherpa.org/wp-content/uploads/2020/03/COVID-19-Student-Privacy-FAQs-03-20-2020-1.pdf; "Thermal Imaging as Pandemic Exit Strategy" (June 3, 2020), https://fpf.org/2020/06/03/thermal-imaging-as-pandemic-exit-strategy-limitations-use-cases-and-privacy-implications/; "COVID-19 and Data Protection Resources," https://sites.google.com/fpf.org/covid-19-privacy-resources, "Artificial Intelligence and the COVID-19 Pandemic" (May 7, 2020). https://fpf.org/2020/05/07/artificial-intelligence-and-the-covid-19-pandemic/

risks. Promising approaches include: decentralized contact tracing, homomorphic encryption, and differential privacy;

- Support the development of emerging privacy-enhancing technologies that hold promise in the public health sphere. Emerging technologies include: synthetic data, controlled access environments, digital twins, and simulations;
- Support cross-disciplinary research into privacy-protective approaches to key emerging technologies, including Wireless Sensor Networks and data processing strategies for on-device and/or centralized analysis of personal health information;
- Explore mechanisms to balance the need for increased access to data that allows researchers to understand the differential impacts of crises on certain communities by not obscuring critical community characteristics with privacy-enhancing technologies;
- Convene cross-disciplinary experts to create and refine guidance for implementation of privacy protections suited to crisis situations;
- Identify top-priority updates to laws and regulations pertaining to public health;
- Explore mechanisms that promote data interoperability while promoting privacy;
- Promote the development of promising de-identification technologies and mitigation strategies to address re-identification risks;
- Promote practical, implementable ethical frameworks that go beyond the FAIR principles; and
- Identify practical lessons learned during the COVID-19 pandemic regarding publication ethics and norms for research in a time of crisis that can apply to future crises.

2.0 Project and Visioning Outline

The NSF Convergence Accelerator program is intended to fund research projects that bring together researchers from multiple disciplines, working for university-based, corporate-based, government-based, or civil society-based research groups. We convened a workshop of stakeholders from these groups to consider the implications of the COVID-19 pandemic in terms of data access, regulatory and policy challenges to researching or building privacy-preserving technologies for public health response, technical challenges to building privacy-sensitive technologies for public health response, and ethical data uses in the setting of COVID-19 research and response.

The participants were invited to submit “position statements,” which were reviewed for quality and interest. Those selected were asked to participate in a two-day workshop addressing four themes:

- Accessibility of Data to Track SARS CoV-2
- Use of Technology to Track, Trace & Notify to Control Spread of COVID-19
- Adapting Legal and Regulatory Responses to a Global Emergency
- The Convergence of Technology, Policy and Responsible Data Use in a Global Crisis

In the summary and analysis that follows, we extend our workshop's findings, identifying areas for further research on public health surveillance technologies, privacy, data access and data use, and implementable data ethics.

3.0 Privacy Law, Regulation, and Technologies

The global reach of the COVID-19 pandemic highlights the importance of national legal frameworks for the protection of personal data. The pandemic also highlights that effective and enforceable privacy legislation must bring together the expertise of both legal and engineering scholars. Many of the effective protections will come from careful research into technological options designed to lead best practices, regulatory guidance, and comprehensive legislation.

3.1 New Directions for Privacy Enhancing and Privacy-Preserving Technologies

Privacy legislation that is sufficiently informed by the degree to which distinct privacy technologies protect individuals from re-identification harms and can establish a risk-adjusted prioritization for uses of these technologies in relevant domains, such as healthcare, represents a goal for meaningful guidance on responsible data uses in a crisis. Privacy-preserving machine learning, differential privacy, data synthesis, digital twins, and repeatable privacy risk assessments are techniques that could be codified into a comprehensive privacy strategy that surpasses the present patchwork of solutions. Identifying these privacy-enhancing technologies (PETs) will come through the convergence of multiple research streams.

3.1.1 Privacy-Preserving Machine Learning

Privacy-Preserving Machine Learning is a combination of hardware and data flow management techniques that allow multiple parties to collaboratively train machine learning models without risking privacy through the transfer of data.[5] Data decentralization, federated "on-device" learning, or a distributed data and hardware approach are all effective techniques to prevent severe privacy breaches such as membership inversion or dataset reconstruction attacks. Selecting criteria for when and how to employ each is an important challenge to address when designing privacy legislation that is sensitive to technical challenges.[6]

3.1.2 Differential Privacy

Differential privacy approaches require the systematic introduction of randomized modifications to a dataset or algorithm to reduce the available amount of information about a single individual

[5] "FPF Webinar Explores the Future of Privacy-Preserving Machine Learning" (July 1, 2020), https://fpf.org/2020/07/01/fpf-webinar-explores-the-future-of-privacy-preserving-machine-learning/.

[6] Loshin, David and Butler, Brian. 2020. "Understanding Organizational Dynamics of Federated Data Collection and Privacy Preservation." Privacy & Pandemics Workshop Working Paper 45. Future of Privacy Forum, October 27; Reina, G. Anthony and Shah, Prashant. 2020. "COVID-19: Federated Learning for Privacy Preserving Multi-Institutional Collaboration." Privacy & Pandemics Workshop Working Paper 6. Future of Privacy Forum, October 27; Kaissis, G.A., Makowski, M.R., Rückert, D. *et al.* Secure, Privacy-Preserving and Federated Machine Learning in Medical Imaging. *Nature Machine Intelligence* 2, 305–311 (2020). https://doi.org/10.1038/s42256-020-0186-1.

without interfering with the integrity of that dataset for analysis. Differential privacy techniques can be applied to the input, analysis, or the output of machine learning models. Determination of which forms of differential privacy transformation (e.g., Laplacian, or Gaussian noise) should be applied at which stage, or for which types of analytical processes is a critical research question, requiring the intersection of both basic and applied statistics and data management research streams.[7]

3.1.3 Homomorphic Encryption

Preserving privacy to the maximum extent technically feasible is a goal particularly when dealing with sensitive data, such as health data gathered about individuals affected by COVID-19. By allowing the analysis and manipulation of data without decrypting it, homomorphic encryption retains strong privacy protections, particularly applicable to sensitive data. However, the computational and monetary costs can be high, and the practical limitations of fully homomorphic encryption present barriers to the use of this privacy-preserving technique for some of the data generated, such as the notes in electronic health records.[8] Research to determine which of the many forms of encryption algorithms limits computational costs while maximizing data privacy is necessary. Likewise, applied computer science research that extends the usefulness of this technique to other forms of data is a clear requirement, highlighted during this pandemic.[9]

3.1.4 Synthetic Data

Synthetic data is data produced to mimic the relevant distributions and other salient characteristics of a sensitive data set but without the actual inclusion of the sensitive data.[10] Synthetic data can be readily produced using most packages used by data scientists and machine learning professionals, but the ease of production masks the complexity of measuring the risk of re-identification or mistaken identification of persons whose real data patterns match those in a synthetic environment.[11] Despite some risk, synthetic data can also be built to augment data sets with few cases or data with quality challenges that make analytics of the original data at scale a challenge. To what extent synthetic data can be built with the lowest re-identification risks is an important research question for computer scientists and privacy engineers to tackle.

[7] Blocki, Jeremiah, Avrim Blum, Anupam Datta, and Or Sheffet. "The johnson-lindenstrauss transform itself preserves differential privacy." In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pp. 410-419. IEEE, 2012, p. 410-411.

[8] Wu, David, and Jacob Haven. "Using homomorphic encryption for large scale statistical analysis." *FHE-SI-Report, Univ. Stanford, Stanford, CA, USA, Tech. Rep. TR-dwu4* (2012).

[9] Parmar, Payal V., Shraddha B. Padhar, Shafika N. Patel, Niyatee I. Bhatt, and Rutvij H. Jhaveri. "Survey of various homomorphic encryption algorithms and schemes." *International Journal of Computer Applications* 91, no. 8 (2014).

[10] El Emam, Khaled and Sood, Harpreet. 2020. "Enabling COVID-19 Data Access Using Data Synthesis." Privacy & Pandemics Workshop Working Paper 56. Future of Privacy Forum, October 27; *Accelerating AI with Synthetic Data* (Apr. 9, 2020). This is an FPF introductory course, co-hosted by Dr. Khaled El Emam of Replica Analytics.

[11] Abowd J.M., Vilhuber L. "How Protective Are Synthetic Data?." In *Domingo-Ferrer J., Saygın Y. (eds) Privacy in Statistical Databases* (2008); Lecture Notes in Computer Science, vol 5262. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-87471-3_20.

3.1.5 Digital Twins

Digital twins, also called data doubles, are digital prototypes or representations of physical systems or processes built upon historical data.[12] Digital twins represent complex physical systems, such as critically ill patients, computationally allowing for low-risk, low-cost, digital testing of drugs, devices, or techniques.[13] The utility of reproducing such systems in digital forms allows researchers to shift risks of real-world testing so they may run countless hypothetical scenarios on data alone, including scenarios that would otherwise be expensive, risky, or even ethically troublesome. Digital twins may also be combined into ensembles, such as is done in smart manufacturing, to represent complex, community health dynamics for testing of novel technical, pharmaceutical, or social solutions.[14]

The privacy implications of using digital twins in health research settings are not well characterized, although there have been concerns analogous to the privacy implications of genomic medicine.[15] Researching the risks of re-identification of persons in a healthcare digital twin setting and risks of re-identification of persons through digital twin ensembles in an internet of things or twinned manufacturing environment represents a novel area of practical privacy research.

3.1.6 Simulations and Modeling

The COVID-19 pandemic highlights the complex interactions and dependencies between social behaviors, viral strains and mutations, testing procedures, population health dynamics, medical technology, global organization, and climate change. Alone, these complex systems represent serious challenges to the model, much less estimating the effects of individual changes on the system. Taken together, the complex dynamics of a pandemic spread represent hundreds or even thousands of parameters to estimate in a model of spread and control. Creating robust simulations and models of the dynamics of individual or system-level interactions, such as through agent-based modeling, represents an avenue for estimating the effects of changes before policies are put into place.[16] Simulations, trade-off results, and models for simulations can guide researchers and policy makers striving to estimate the risks and benefits associated with a change in a complex system, such as pandemic health dynamics.

[12] Digital Twin Consortium, “Frequently Asked Questions,” https://www.digitaltwinconsortium.org/faq.htm.

[13] Björnsson, Bergthor., Borrebaeck, Carl., Elander, Nils. *et al.* “Digital twins to personalize medicine.” *Genome Medicine* 12, 4 (2020). https://doi.org/10.1186/s13073-019-0701-3

[14] A. Fuller, Z. Fan, C. Day and C. Barlow, "Digital Twin: Enabling Technologies, Challenges and Open Research," in *IEEE Access*, vol. 8 (2020), pp. 108952-108971, doi: 10.1109/ACCESS.2020.2998358.

[15] Bruynseels, Koen, Filippo Santoni de Sio, and Jeroen van den Hoven. "Digital twins in health care: ethical implications of an emerging engineering paradigm." *Frontiers in genetics* 9 (2018): 31.

[16] DeCesare, Ana, et al. 2020. "Trade-Off Between Privacy and Efficacy." Privacy & Pandemics Workshop Working Paper 54. Future of Privacy Forum, October 27.; Cuevas, Erik. “An agent-based model to evaluate the COVID-19 transmission risks in facilities.” *Computers in biology and medicine* vol. 121 (2020): 103827. doi:10.1016/j.compbiomed.2020.103827

3.1.7 Guidance for implementation of PETs during Crises

All crises are different; just as a hurricane in one time or place is different from a hurricane in another time or place, the data gathered and its uses during a crisis response are not the same. The high degree of contextual variability for crises creates challenges for establishing guidelines for *ex-ante* implementation of PETs. However, this challenge should not stand in the way of a development of clear guidance for the implementation of specific PETs to manage data flows during even the earliest days of a crisis, given known common dimensions of crises, such as flooding, wind, or fire disasters.[17]

3.2 Adapting Privacy Law Research for New Technologies

The COVID-19 pandemic raises the possibility that major tenets of public health law must be revisited to account for the emergence of technological approaches to address public health problems. Law and policy will have to address contact tracing, exposure notification, and other novel technology applications to public health.

3.2.1 Public Health Surveillance and Privacy

The pandemic crisis reveals that ensuring data protections across the complex interdependence of public and personal health systems is not easily accomplished under current laws and regulations.[18] Articulation of clear guidance for preserving privacy for the many forms and uses of personal health information will be a persistent need throughout and after the COVID-19 pandemic. Concerns about the privacy and security of health data motivates many questions about the applicability of the Health Information Portability and Accountability Act (HIPAA) Privacy Rule and the Common Rule (45CFR46),[19] including whether these two laws adequately govern the conduct of public health research and surveillance initiatives involving new technologies, novel computational solutions, and innovations produced by non-traditional industry stakeholders. Confusion regarding the definitions of permitted versus barred uses of healthcare data, as defined in existing legal and regulatory frameworks, can cause delays or missteps by actors within and outside of traditional health research and public health surveillance contexts. For example: 1) whether track and trace applications data could be used by researchers to answer questions about individuals' movements and transportation needs in the pandemic; or 2) whether contact tracing data (digital or otherwise) could be used by law enforcement to monitor public political behavior emerged as consequential questions without clear answers. The lack of legal and regulatory clarity has complicated efforts to manage and protect data in cross-

[17] S. Parasarathay. Verbal Presentation. Privacy & Pandemics. October 28, 2020. Virtual Conference.

[18] Washington, Anne L. and Rhue, Lauren. 2020. "Interlocking Decision Systems and Disparity." Privacy & Pandemics Workshop Working Paper 31. Future of Privacy Forum, October 27.

[19] Specifically, 45CFR46.102(l)(2) and 45CFR46.102(k) and related discussion during the October 20, 2020 discussion by the Secretaries Advisory Committee on Human Research Protections (SACHRP).

disciplinary collaborations to address the pandemic, and has impacted public trust is such systems used for either purpose.[20]

3.2.2 Contact Tracing and Exposure Notification Technologies

Applications designed to perform public health surveillance functions previously carried out by professionals in an “analog” environment, such as manual contact tracing, have changed the types of public health Information collected, stored, and shared. The most widely adopted apps in the US are deployed by public health authorities leveraging Bluetooth proximity for exposure notification in a decentralized privacy-preserving manner. Contact tracing apps, which may rely on geolocation and store collected data in a central server, can also be deployed by schools, employers, and private companies as part of their comprehensive strategies to manage students, employees, and clients returning to their physical premises during the pandemic. Exposure notification and contact tracing apps may include additional functionalities, such as symptom-trackers, quarantine management, and the provision of vital information. Depending on the design of an app, different risks arise involving not only privacy, but also accuracy, precision, transparency, and equity.

The pandemic underscores the vital role of public trust in the adoption of socially beneficial technologies.[21] For instance, now, contact tracing technologies and apps require voluntary adoption and active participation by the public in order to draw meaningful conclusions about the technologies' or apps' ability to engender positive public health outcomes.[22] However, in the US, adoption of contact tracing apps has been undermined by a seeming lack of trust and a perception that data will be collected, used, shared, and retained in inappropriate ways by both public and private entities.[23] This chilling effect, caused in part by a lack of adequate data privacy legislation in the US, contrasts with relatively higher adoption rates in European jurisdictions, such as Ireland, Germany, and Switzerland, which enjoy the protections afforded by the General Data Protection Regulation (GDPR).[24] Recognizing this perception, a patchwork of federal and state legislation has been introduced in the US to protect the confidentiality of digital contact tracing app data and other public health emergency data. Sharing limitations are common features of many proposals, with concerns about sharing emergency health data with government

[20] Fry, Caroline V., Xiaojing Cai, Yi Zhang, and Caroline S. Wagner. "Consolidation in a crisis: Patterns of international collaboration in early COVID-19 research." *PloS one* 15, no. 7 (2020): e0236307.

[21] Simpson, Erin and Collins, Sara. 2020. "Trust Deficit Why a Lack of Trust in Government and Technology has Harmed our Pandemic Response." Privacy & Pandemics Workshop Working Paper 43. Future of Privacy Forum, October 27; Unger, Wayne. 2020. "Katz and COVID-19: How a Pandemic Changed the Reasonable Expectation of Privacy." Privacy & Pandemics Workshop Working Paper 22. Future of Privacy Forum, October 27.

[22] Weissinger, Laurin B. 2020. "Tech is not the Limit | Trust is: Why Apps Could Not Solve this Crisis and Will Not Solve the Next." Privacy & Pandemics Workshop Working Paper 51. Future of Privacy Forum, October 27.

[23] Kissick, Chas, et al. 2020. "Evaluating Contact Tracing Apps: We Need More Transparency." Privacy & Pandemics Workshop Working Paper 46. Future of Privacy Forum, October 27.

[24] Future of Privacy Forum. “Why Data Protection Law is Uniquely Equipped to Let Us Fight a Pandemic with Personal Data”. (April 27, 2020). Available at: https://fpf.org/2020/04/07/why-data-protection-law-is-uniquely-equipped-to-let-us-fight-a-pandemic-with-personal-data

agencies, such as law enforcement. Many bills in the US have also sought to introduce requirements relating to retention and purpose limitations, data security, transparency requirements about data practices to individuals, and to obtain individuals' revocable consent for data collection and usage. Transparency, in particular, is a fundamental trust-based incentive to promote accountability and responsible data use.[25]

However, at a fundamental level, data-driven technologies that have emerged to mitigate the spread of COVID-19 -- including not only contact tracing and exposure notification apps, but also AI/ML-powered automated decision-making affecting large groups -- demonstrate the urgent need for regulatory frameworks to account for the risks not only   involving individual privacy invasions caused by the misuse or abuse of personal information, but also the broader social risks, such as those associated with exposure of individuals' social networks. This urgency may lead to discrimination, exclusion, and inequality. Data protection frameworks like the GDPR are well-equipped to apply in these contexts. At a broader societal level, data protection regulatory models, such as the GDPR, require data protection by design (Art.25), data protection impact assessments for large scale or sensitive data processing (Art.35), and all processing of personal data must be fair and transparent (Art.5(1)(a)) - regardless of consent. Ensuring that such regulatory coverage extends to more jurisdictions could be a key legislative and regulatory change to emerge from the pandemic.

3.2.3 Comparative Effectiveness of Data Protection Regimes

The COVID-19 pandemic highlights strong variation in the implementation flexibility of national and subnational approaches to data protection and the control of personal data flows.[26] Thus, the pandemic has elevated three questions of importance to the governance of data and technology: 1) do strong, centralized, legal and regulatory data protection regimes create more or less genial environments for technological innovation; 2) do strong, centralized, legal and regulatory data protection regimes cause businesses to incur an outsized cost for privacy compliance when compared to fragmented or more decentralized regimes; and 3) do strong, centralized, legal and regulatory data protection regimes create more genial conditions for the protection of personal data at the individual level? A convergent research program, bringing together data science, data protection, comparative politics, comparative economics, and business management experts, could help answer these questions.

4.0 Data Access and Data Use: Practical Challenges

Crises surface problems and solutions that could not have been foreseen. The COVID-19 pandemic shows that the challenge to the coordination of data sources is a problem that can be

[25] Listokin, Siona. 2020. "Priceless or Worthless? Formal Privacy Valuation in COVID-19 Policy Proposals." Privacy & Pandemics Workshop Working Paper 37. Future of Privacy Forum, October 27.

[26] Dean Parker, Kimball and Kulbeth, Marie. 2020. "Flexible Regulation in the Time of COVID." Privacy & Pandemics Workshop Working Paper 12. Future of Privacy Forum, October 27.

solved by public and private actors.[27] The pandemic also demonstrates that data quality remains a persistent challenge for research and application.

4.1 Practical Challenges to Data Access

Data accessibility is limited by control of data assets by single entities, lack of interoperability, and multiple data quality problems. Initiatives to increase access to “real world evidence” and “real world data” have succeeded in some areas and fallen short in others. The COVID-19 pandemic demonstrated that fixing data accessibility is possible when capital and institutional will are devoted to the cause.

4.1.1 Coordination of Data and Data Access

During the early months of the COVID-19 pandemic, data analysis and artificial intelligence experts determined that the potential of analytics to help organizations make critical decisions was limited so long as data remained siloed in discrete locations.[28] Efforts by governments to create centralized repositories were met, and even eclipsed, by similar efforts by private companies, civil society organizations, and motivated individuals.[29] These many efforts demonstrated that the challenges of creating centralized data repositories, such as those described by the National Academies and others, were tractable problems.[30] Improvements to pandemic-related data access are only one part of the cross-industry, multidisciplinary efforts that are needed to create accessible data for truly novel analytics at scale. Further research initiatives that identify barriers to building accessible data may not be needed, but other resources such as funding, personnel, governance, and effective data architecture to speed data access are.[31]

4.1.2 Interoperability as Access Problem

Interoperable data is hailed as a key to improving healthcare and industrial applications. However, as multiple authors have pointed out, “most of today’s medical data lack interoperability: hidden in isolated databases, incompatible systems and proprietary software, the data are difficult to exchange, analyze, and interpret. This slows down medical progress, as technologies that rely on these data—artificial intelligence, big data or mobile applications—

[27] Blackport, Jamie, Moffatt, Colin, Kassam-Adams, Shahir, and Brennan, Niall. 2020. "COVID-19 Research Database: A Case Study of Pragmatic Patient Privacy Protection." Privacy & Pandemics Workshop Working Paper 14. Future of Privacy Forum, October 27.

[28] Greenbaum, Dov, Gursoy, Gamze, and Gerstein, Mark. 2020. "Making Real-World Data Useful within the New Drug Application Process." Privacy & Pandemics Workshop Working Paper 25. Future of Privacy Forum, October 27.

[29] National Institutes of Health, Office of Data Science Strategy. “Open-Access Data and Computational Resources to Address COVID-19,” https://datascience.nih.gov/covid-19-open-access-resources; Mathematica.org. “COVID-19 Curated Data, Modeling and Policy Resources,” https://www.mathematica.org/features/covid-19-curated-data-modeling-and-policy-resources; Harvard University, Harvard Dataverse. “COVID-19 Data Collection,” https://dataverse.harvard.edu/dataverse/covid19; C3.ai. “C3 AI COVID-19 Data Lake,” https://c3.ai/products/c3-ai-covid-19-data-lake/.

[30] Abhishek Nagaraj, Esther Shears, Mathijs de Vaan. “Improving data access democratizes and diversifies science.” *Proceedings of the National Academy of Sciences* 117, no. 38 (2020): 23490-23498, DOI: 10.1073/pnas.2001682117; National Research Council. 2005. “Expanding Access to Research Data: Reconciling Risks and Opportunities.” Washington, DC.

[31] Parthasarathy, Srinivasan. 2020. "Data Science Technology and Governance Challenges During Crisis Response." Privacy & Pandemics Workshop Working Paper 39. Future of Privacy Forum, October 27.

cannot be used to their full potential."[32] Calls for data interoperability are often too general, neglecting the types of interoperability needed to facilitate large scale analytics: technical, syntactic, semantic, and organizational interoperability are all necessary components of a fully interoperable data environment. Adopting technical interoperability standards across sensor networks will present engineering challenges soon, but overcoming these challenges may be undone by inattention to syntactic and semantic interoperability failures. For the healthcare setting, in particular, FHIR (Health Level 7's Fast Healthcare Interoperability Resources) and SNOMED CT represent advances in syntactic and semantic interoperability that could be replicated in other domains.[33] Supporting research that examines the cost savings and analytics acceleration from these standards may help to accelerate similar interoperability components for adjacent industries.

4.1.3 Standardized Data Quality as Problem for Data Accessibility

The COVID-19 pandemic demonstrated that data can be made accessible, given resources and intention, but that the quality of that data is what stands between access and analysis. Projects such as the American Heart Association's "Get With the Guidelines" (GWTG) registry helped to speed useful data to the hands of researchers.[34] However, in the absence of clear data standards, healthcare access and quality, and consistent reporting guidelines, there was a loss of essential information about individual and community clusters.[35] The consequences of missing, noisy, or inconsistent data come into stark relief in applying large scale analytics to low-quality data.[36] Even promising systems falter when data quality issues introduce an intolerable level of uncertainty to measurements of outcomes for decision-making. While there is a steady stream of research on the imputation of missing values and large scale analytics using data sets with missing values, this research is often theoretical, and the consequences of the use of the proposed techniques for real world analytics are not wholly characterized.[37] Estimating the cost of gaps in data access and data quality represents a research problem salient to virtually all research areas funded by the NSF.

4.1.4 Data Comprehensibility that Builds Trust as Leading to Data Accessibility Problems

Accessible data comes from many sources, some of which are individuals who decide to share their data with organizations that they trust. Data trusts are one mechanism that allow

[32] Lehne, M., Sass, J., Essenwanger, A. *et al.* "Why digital medicine depends on interoperability." *npj Digit. Med.* 2, 79 (2019), https://doi.org/10.1038/s41746-019-0158-1.

[33] HL7 International. (2019). "Overview- FHIR v4.0.1," https://www.hl7.org/fhir/overview.html; SNOMED International. (2020). "SNOMED International," http://www.snomed.org/.

[34] Alger, Heather M., Joseph H. Williams IV, Jason G. Walchok, Michele Bolles, Gregg C. Fonarow, and Christine Rutan. "Role of Data Registries in the Time of COVID-19." *Circulation: Cardiovascular Quality and Outcomes* 13, no. 5 (2020): e006766.

[35] Lau, Hien, Veria Khosrawipour, Piotr Kocbach, Agata Mikolajczyk, Hirohito Ichii, Justyna Schubert, Jacek Bania, and Tanja Khosrawipour. "Internationally lost COVID-19 cases." *Journal of Microbiology, Immunology and Infection* (2020).

[36] Cortes, Corinna, Lawrence D. Jackel, and Wan-Ping Chiang. "Limits on learning machine accuracy imposed by data quality." In *Advances in Neural Information Processing Systems*, pp. 239-246. 1995.

[37] Batista, Gustavo EAPA, and Maria Carolina Monard. "An analysis of four missing data treatment methods for supervised learning." *Applied artificial intelligence* 17, no. 5-6 (2003): 519-533.

individuals to share their data to a central location that then vets the onward transfer of data for specific purposes.[38] Outside of well-coordinated, purpose-driven efforts, such as data trusts, individuals may struggle to identify whether sharing their personal data fits their individual or group-membership values. Without effective mechanisms to explain data transparently so as to increase individuals' comprehension of the uses of their data, the trust will lag, stymying efforts.

Two mechanisms for increasing trust in the uses of data are "data nutrition labels" and "datasheets for datasets." "Data nutrition labels" are "designed to support specific decisions by the consumer," allowing them to participate as an effective stakeholder in data sharing exercises.[39] These labels help to equalize the power differential between those whose lives the data captures and those who capture data about others' lives. Another is the ongoing project to draft "datasheets for datasets." Public knowledge of these efforts and the ability of members of the public to meaningfully use this information to make informed decisions is an open question to be answered through careful social science and education research.

4.2 Practical Challenges to Data Use

Which methods for protecting privacy are most efficacious to protect persons while also presenting the most refined levels of useful information for analysis? Whether data centralization presents a privacy advantage vis-a-vis decentralization is not a settled question, whether theoretically or in specific applications. Likewise, it is not clear which methods of de-identification, aggregation, or differentiation reduce re-identification and privacy risks to preferred levels. Creating metrics for evaluating the true risks and the costs of the potential harms from the re-identification of data is an area ripe for cross-cutting research. Finally, assessing the analytic loss for the use of any of these methods has not been calculated.[40]

4.2.1 Data de-identification

Data de-identification is one approach to the challenge of preserving privacy during data use.[41] However, any use of data presents some risk of re-identification. Efforts to measure those risks within specific industry verticals or as pertains to specific types of data appear in the academic and industry-produced literature.[42] Yet, research on true re-identification risks has not grown in

[38] The Open Data Initiative. 2019. "Data Trusts: Legal and Governance Considerations," https://theodi.org/article/data-trusts-legal-report/; Future of Privacy Forum. (2020). "FPF & Brighthive Release Playbook," https://fpf.org/2020/07/21/fpf-brighthive-release-playbook-to-create-responsible-contact-tracing-initiatives-address-privacy-ethics-concerns/.

[39] Stoyanovich, Julia, and Bill Howe. "Nutritional Labels for Data and Models." *IEEE Data Eng. Bull.* 42, no. 3 (2019): 13-23. P. 13-14.

[40] Xu, Heng, and Nan Zhang. "Implications of Data Anonymization on the Statistical Evidence of Disparity." *Available at SSRN 3662612* (2020); Xu, Heng, and Nan Zhang. "Privacy in Health Disparity Research." *Medical care* 57 (2019): S172-S175.

[41] Future of Privacy Forum, "FPF: A Visual Guide to Practical Data De-Identification" (April 25, 2016), https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/.

[42] McGuinness, Niamh and Arbuckle, Luk. 2020. "Challenges in Access to Patient-Level COVID-19 Data and the Role of De-Identification." Privacy & Pandemics Workshop Working Paper 16. Future of Privacy Forum, October 27; El Emam, Khaled, Luk Arbuckle, Gunes Koru, Benjamin Eze, Lisa Gaudette, Emilio Neri, Sean Rose, Jeremy Howard, and Jonathan Gluck. "De-identification methods for open health data: the case of the Heritage Health Prize claims dataset." *Journal of medical Internet research* 14, no. 1 (2012): e33; Nelson, Gregory S. "Practical implications of sharing data: a primer on data privacy, anonymization, and de-identification." In *SAS Global Forum Proceedings*, pp. 1-23. 2015; Raisaro, Jean Louis, Florian Tramer,

proportion to the number of fields now routinely using advanced analytics and de-identified data. Support of cross-cutting research that enables practical de-identification of data and for estimating the risk of re-identification is an area in need of attention.

Re-identification risks are not the only hazards that arise from efforts to manage data through de-identification. Emerging research shows that de-identification techniques may obscure valuable information, such as demographic and socioeconomic status information, necessary to evaluate all dimensions of need in healthcare and other data-driven research.[43] Associated research asks whether the risks of re-identification that arise from uses of location data, such as that used in public health surveillance technologies, might eclipse the utility of that data.[44] A coordinated and multidisciplinary research program that evaluates the risks and benefits that arise when data is de-identified, and which evaluates those risks with respect to the research objectives of specific fields, the preferences of specific groups, and novel alternatives (e.g., synthetic data) is necessary.

Concerns about re-identification risks may keep some researchers from pursuing valuable research streams, whether due to institutional actors restricting the pursuit of the research or due to researchers' self-policing.[45] The sociology of data science practices around collaborative data uses, including de-identification, presents an avenue of important research on the norms of analytic science.[46]

4.2.2 Data Sufficiency

The early days of the COVID-19 pandemic showed how access to only a small number of time-limited data points can hinder the development of essential analytics and intelligence. With sufficient support, research into low data use machine learning may solve this problem for other crises where limited data flows hinder decision analytics.[47] However, the pandemic also showed how the problem of data sufficiency or whether an incompletely collected (because still emerging) data time series will be enough to support decision-making in a high-consequence scenario.[48] In the situation of an evolving crisis, when data unfolds as a stream, and where all

Zhanglong Ji, Diyue Bu, Yongan Zhao, Knox Carey, David Lloyd *et al.* "Addressing Beacon re-identification attacks: quantification and mitigation of privacy risks." *Journal of the American Medical Informatics Association* 24, no. 4 (2017): 799-805.

[43] Xu, Heng, and Nan Zhang. "Implications of Data Anonymization on the Statistical Evidence of Disparity." *Available at SSRN 3662612* (2020); Xu, Heng, and Nan Zhang. "Privacy in Health Disparity Research." *Medical care* 57 (2019): S172-S175.

[44] Yin, Ling, Qian Wang, Shih-Lung Shaw, Zhixiang Fang, Jinxing Hu, Ye Tao, and Wei Wang. "Re-identification risk versus data utility for aggregated mobility research using mobile phone location data." *PloS one* 10, no. 10 (2015): e0140589.

[45] Stalla-Bourdillon, Sophie, et al. 2020. "Empowering Researchers in Times of Crisis." Privacy & Pandemics Workshop Working Paper 50. Future of Privacy Forum, October 27.

[46] Barth-Jones, Daniel. "The're-identification'of Governor William Weld's medical information: a critical re-examination of health data identification risks and privacy protections, then and now." *Then and Now (July 2012)* (2012).

[47] Xian, Yongqin, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. "Feature generating networks for zero-shot learning." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5542-5551. 2018.

[48] Ahn, G.; Lee, H.; Park, J.; Hur, S. Development of Indicator of Data Sufficiency for Feature-based Early Time Series Classification with Applications of Bearing Fault Diagnosis. *Processes* 2020, *8*, 790.

data is treated as equally valuable within that stream as an *a priori* condition, there are poorly-characterized risks to the uses of sophisticated analytics.[49] Similarly, the privacy risks that may arise when potentially identifying information is repeated within a stream, whether as a function of repetition over time or data sources, are not well characterized. Determining the risks to analytics and privacy of high-velocity data analytics in a crisis represents a genuine gap for convergent teams to fill.

4.2.3 Data Aggregation

Aggregate data presents fewer re-identification risks to individuals or groups than does higher resolution data.[50] However, reporting of aggregate data may present unintended risks when the number of cases in a geographical catchment is small.[51] Data aggregation techniques may also obfuscate important dimensions of diversity, such as the collapse of race into a binary variable.[52] For some types of data, however, aggregation may present a better picture of true risk to an individual, such as when exposure events are aggregated over a period to represent a cumulative risk score.[53]

4.3 Practical Challenges to Data Sharing

Data sharing is a laudable exercise laden with practical challenges.[54] Current efforts to share data across organizations, particularly from corporate organizations to research teams,[55] are often restricted to organizations seeking to share data for specific purposes using bespoke tools, such as data use contracts. The challenges of data sharing for the purpose of developing machine learning research and tools are compounded by the uncertainty of both data sharing and data use for machine learning and artificial intelligence.

To effectively manage the risks and benefits of data sharing, equal attention must be paid to the probability and magnitude of risks attached to all components of a research data sharing exercise, including assessment of hardware and software, data collection, data privacy, and models.[56]

[49] Charu C. Aggarwal. 2003. A framework for diagnosing changes in evolving data streams. In Proceedings of the 2003 ACM SIGMOD international conference on Management of data (SIGMOD '03). Association for Computing Machinery, New York, NY, USA, 575–586. DOI:https://doi.org/10.1145/872757.872826; Wei Fan. 2004. StreamMiner: a classifier ensemble-based engine to mine concept-drifting data streams. In Proceedings of the Thirtieth international conference on Very large data bases - Volume 30 (VLDB '04). VLDB Endowment, 1257–1260.

[50] Future of Privacy Forum. (2016). "*FPF: A Visual Guide to Practical Data De-Identification"* Available at: https://fpf.org/2016/04/25/a-visual-guide-to-practical-data-de-identification/

[51] Katz, Aaron. Workshop Presentation. Privacy & Pandemics. October 27, 2020.

[52] Washington, Anne L., et al. 2020. "Categories of COVID." Privacy & Pandemics Workshop Working Paper 49. Future of Privacy Forum, October 27.

[53] Bater, Johes, et al. 2020. "Poirot: Private Contact Summary Aggregation." Privacy & Pandemics Workshop Working Paper 27. Future of Privacy Forum, October 27.

[54] Future of Privacy Forum. 2020. *Best Practices for Sharing Data With Academic Researchers*. , https://fpf.org/2020/10/28/fpf-best-practices-and-contract-guidelines-help-companies-share-data-with-academic-researchers/

[55] Future of Privacy Forum. (2015). "*Beyond IRBs: Designing Ethical Review Processes for Big Data Research: Conference Proceedings*". Available at: https://fpf.org/wp-content/uploads/2017/01/Beyond-IRBs-Conference-Proceedings_12-20-16.pdf

[56] Jordan, Sara R. (2019). "Designing an Artificial Intelligence Research Review Committee". Available at: https://fpf.org/wp-content/uploads/2019/10/DesigningAIResearchReviewCommittee.pdf; Čerka, P., Grigienė, J., & Sirbikytė, G. (2015). *Liability*

4.3.1 Issues in Data Sharing: Data Risk, Model Risks, and Privacy Risk Assessments

Credible and legitimate governance of sharing exercises must be established to build confidence in data sharing. Risk assessments are processes that help organizations estimate their tolerance for risk to themselves and to the individuals who have contributed data through the use of their products and services. Multiple risk estimation methodologies, risk models, and approaches to determining salient risk factors complicate the implementation of reproducible risk assessments.[57] Within the privacy impact assessment arena, these complicating factors lead to a lack of repeatability, whether within organizations or across product or process types.[58] Research that robustly classifies risk assessment parameters, estimates risks and benefits across domains, solidifies statistical methods for effective, repeatable, and durable risk impact assessments is essential to moving forward to implement risk assessment in data sharing exercises.

Specific characteristics of datasets introduce distinct risks or benefits. These characteristics include data quality and data quantity, data provenance, association of data sets with legal authorization for their use, level of data refinement or data aggregation, and whether the data is held and processed by firms in specific locations (e.g., the EU) or specific industry verticals (e.g., data brokers).[59] Data characteristics also include the mechanism by which data is collected, such as IoT data, images, CCTV, vehicle data, or category of data described in specific laws, such as health data, mobility and transportation data, and education data.[60] Which elements of a data-centered approach are most salient for researchers and research managers' descriptions of risk are not yet known. Likewise, the performance gains from using more beneficent and less risky data are not yet known. Research teams must be convened across the many disciplines to establish salient metrics for assessment of data-specific risks and to translate those metrics into useful assessment tools.

Estimations of the risks of data sharing should not focus solely on the data. As specific forms of a model and particular types of model architecture are more susceptible to attack, failure to conduct due diligence checks on model security represents a type of risk introduced outside of the concerns associated with data as such. Critical questions must be addressed, such as: does the use of a complex algorithm introduce risks that are attributable to the model itself; and does

for damages caused by artificial intelligence. Computer Law & Security Review, 31(3), 376-389; Kanal, L. N., & Lemmer, J. F. (Eds.). (2014). *Uncertainty in artificial intelligence*. Elsevier.

[57] https://www.nist.gov/itl/applied-cybersecurity/privacy-engineering/collaboration-space/focus-areas/risk-assessment

[58] Notario, Nicolás, Alberto Crespo, Yod-Samuel Martín, Jose M. Del Alamo, Daniel Le Métayer, Thibaud Antignac, Antonio Kung, Inga Kroener, and David Wright. "PRIPARE: integrating privacy best practices into a privacy engineering methodology." In *2015 IEEE Security and Privacy Workshops*, pp. 151-158. IEEE, 2015.

[59] Cortes, C., Jackel, L. D., & Chiang, W. P. (1995). *Limits on learning machine accuracy imposed by data quality*. In Advances in Neural Information Processing Systems, 239-246.

[60] Center for Machine Learning and Intelligent Systems, University of California Irvine. (2020). UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/index.php; Stanford, S. & Iriondo, R. (2018). *Best Public Datasets for Machine Learning and Data Science*. Towards AI. https://medium.com/towards-artificial-intelligence/the-50-best-public-datasets-for-machine-learning-d80e9f030279; *Open Machine Learning* (2020). Openml.org;

the adoption of XAI (explainable AI) techniques mitigate the risks associated with uses of complex algorithms?[61] Future research must bring together partners from AI/ML research and practice, specifically from the study of interpretable machine learning, explainable AI, privacy-preserving machine learning, and security for machine learning models to examine the risks borne of algorithms alone.[62]

Privacy risk assessments focus on the persons represented in the data and evaluated through the models. Also described as privacy impact assessments, privacy risk assessments are a well-developed component of corporate privacy protection programs. Streamlining privacy impact assessments, incorporating these into research environments, and blending them with data and model risk assessments is a task for applied research investigations that incorporate privacy law, business, research administration, and data management research. Determining how these risk assessments do or do not slow research or protect persons will be another needed stream of multidisciplinary research into data sharing.

4.3.2 Issues in Data Sharing: Sustainable Data Repositories

Crises produce tremendous amounts and varieties of data. Capturing this data for immediate use presents one set of challenges while capturing, cleaning, controlling, moving (when necessary), and storing this data for future use presents another. Research data repositories, whether for single disciplines or multiple disciplines, are a well-established tool for accomplishing these tasks but also have limitations. Research data repositories that merge massive corporate data sets are not as common as corporate data lakes that merge publicly available research data into their environment. Identifying appropriate resource methods for integrating research data repositories and data lakes, which also incorporate appropriate incentives and recognition for the curation of datasets, is a challenge that will arise as pandemic data settles into normalized places.

The establishment of research repositories is not a trivial task. Whether data should be held in corporate-controlled cloud environments, through networks of high-performance computing centers at government-funded laboratories or universities, or through other mechanisms altogether, is an unsettled practical question of technology policy. Additionally, complex is the determination of how stored data could be used under controlled conditions and how this data may be decommissioned or destroyed when it is replaced by a more complete or clean data set; or because the privacy risks associated with its use become too great.

4.3.3 Issues in Data Sharing: Controlled Access Environments

[61] Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). *Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI*. arXiv preprint arXiv:1902.01876.

[62] Holzinger, A. (2018, August). *From machine learning to explainable AI*. In 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA), 55-66. IEEE; Gunning, D. (2017). *Explainable artificial intelligence (xai)*. Defense Advanced Research Projects Agency (DARPA), ND Web, 2, 2.

Novel pathogens will present challenges for socio-technical systems, such as data sharing arrangements. However, tested solutions to the sharing of sensitive data, such as the use of controlled access environments and “Data Access Committees” (DACs) for sharing of high-resolution genomic data (e.g., dbGaP), are systems that can be repurposed for controlled sharing of other sensitive data.[63] Although there remain challenges to ensuring complete protection from the re-identification of individuals through uses of such sensitive data as genomic data, controlled access environments provide governance structures, including oversight and penalties, that reduce re-identification risks to as low as is reasonably achievable.[64]

4.4 *Theoretical Research on Data Access and Data Use*

Applied data analysis rests on sophisticated research in theoretical mathematics, statistics, and computer science. Ensuring that the pipeline of analytics insights remains open through the development of strategies that combine the state-of-the-art level performance on data access, promote privacy-preserving uses, and account for security while ensuring robust analytic performance presents mathematics and computer science researchers with a knotty technical problem ripe for theoretical exploration.[65] Examples include “randomness beacons,”[66] adaptations of differential privacy approaches to novel artificial intelligence algorithms,[67] to edge computing applications generally,[68] and to sensor networks.[69] Modeling challenges to applications of privacy preservation mechanisms, such as data loss and security, and developing practical responses to those challenges, represent an important horizon for computer science, mathematics, and privacy engineering research.[70]

5.0 Data Ethics: Beyond FAIR

“Do no harm” is a basic ethical principle for many professions, including managing and using data. The harms that arise from uses of data are, in many cases, different from the harms that

[63] Shabani M, Dyke SOM, Joly Y, Borry P (2015) Controlled Access under Review: Improving the Governance of Genomic Data Access. PLoS Biol 13(12): e1002339. https://doi.org/10.1371/journal.pbio.1002339;

[64] Church, George, Catherine Heeney, Naomi Hawkins, Jantina de Vries, Paula Boddington, Jane Kaye, Martin Bobrow, Bruce Weir, and P3G Consortium. "Public access to genome-wide data: five views on balancing research with privacy and protection." *PLoS Genet* 5, no. 10 (2009): e1000665.

[65] Mironov, Ilya. "Rényi differential privacy." In *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, pp. 263-275. IEEE, 2017. National Institutes of Standards and Technology. “Privacy-Enhancing Cryptography”. https://csrc.nist.gov/projects/pec

[66] Brandao, Luis. (2020). “Randomness Beacons as Enablers of Public Auditability”. Available at: https://csrc.nist.gov/CSRC/media/Projects/pec/documents/stppa-01-20200127-talk01-brandao-rand-beacons.pdf

[67] Zhao, Jingwen, Yunfang Chen, and Wei Zhang. "Differential privacy preservation in deep learning: Challenges, opportunities and solutions." *IEEE Access* 7 (2019): 48901-48911;

[68] Miao, Qiucheng, Weipeng Jing, and Houbing Song. "Differential privacy–based location privacy enhancing in edge computing." *Concurrency and Computation: Practice and Experience* 31, no. 8 (2019): e4735.

[69] Lin, Chi, Zihao Song, Houbing Song, Yanhong Zhou, Yi Wang, and Guowei Wu. "Differential privacy preserving in big data analytics for connected health." *Journal of medical systems* 40, no. 4 (2016): 97.

[70] Garfinkel, Simson L., John M. Abowd, and Sarah Powazek. "Issues encountered deploying differential privacy." In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, pp. 133-137. 2018.

arise from uses of physical or chemical tools, as in the case of medicine or engineering.[71] However, as has been argued, “Digital harms are harm” and professional obligations to reduce harm extend to the reduction of digital harms.[72] Essential to the reduction of digital harms is the implementation of data ethics principles.

At a minimum, ethical data is data that can be used. The FAIR principles for data—that data is Findable, Accessible, Interoperable, and Reproducible—are *de minimis* conditions for the production of data that can be used to its ethical extent. The FAIR principles must be extended to advance data ethics in the fullest sense of the term, where data ethics includes the social implications of data use, including equity and justice.[73] The addition of any further ethical principles for guiding ethical data use must be motivated by strong evidence of their need and efficacy in practical control of data flows.

5.1 Implementable Data Ethics: Personal Control

Legal frameworks around the world increasingly emphasize the rights of individuals to assert personal control over the acquisition, use, storage, sharing, sale, or destruction of their personal information. Advocates for digital identity suggest that the pandemic has highlighted the need to accelerate mechanisms for individuals to insert themselves as agents to control the flow of data that they create through their interactions with the digital world.[74] Complying with the legal and ethical requirements to recognize these rights presents technical challenges of both a quotidian and novel nature. The quotidian challenge of removing all instances of an individuals’ representation in a dataset can be met with the application of known measures. However, the increased use of individuals’ data to train machine learning systems represents a novel challenge for the removal of the echoes of an individuals’ representation in a dataset.[75] Determination of the value that a single individual brings to a dataset is an increasingly tractable problem,[76] but determining the loss to learning systems and to groups when an individual with unique characteristics is eliminated has not been well characterized.[77] Research into rapid estimation of

[71] Unfairness by Algorithm. (2017). https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/

[72] V. Cassard. Workshop Presentation. “Privacy & Pandemics”. October 28, 2020. Electronic Conference.

[73] Feys, Magali. 2020. "Even Desperate Times Call for 'Fair Trade Data.'" Privacy & Pandemics Workshop Working Paper 47. Future of Privacy Forum, October 27.

[74] Young, Kaliya and Yang, Lucy. 2020. "The COVID-19 Credentials Initiative: Bringing Emerging Privacy-Preserving Technology to a Public Health Crisis." Privacy & Pandemics Workshop Working Paper 10. Future of Privacy Forum, October 27.

[75] Jia, Ruoxi, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gurel, Bo Li, Ce Zhang, Dawn Song, and Costas Spanos. "Towards efficient data valuation based on the shapley value." *arXiv preprint arXiv:1902.10275* (2019).; https://bair.berkeley.edu/blog/2019/12/16/data-worth/

[76] Ghorbani, Amirata, and James Zou. "Data shapley: Equitable valuation of data for machine learning." *arXiv preprint arXiv:1904.02868* (2019).

[77] MacLeod, Haley, Shuo Yang, Kim Oakes, Kay Connelly, and Sriraam Natarajan. "Identifying rare diseases from behavioural data: a machine learning approach." In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pp. 130-139. IEEE, 2016.

the specific value of an individual within a dataset is an essential component of efforts creating a data ecosystem that values individuals as agents.

Knowing the value that a particular individual brings to a dataset does not answer the ethical question of whether any particular individual should have personal control over their inclusion in a dataset. Models of personal control of data flows have been proposed, including personally assertable digital identities, digital doubles, “vendor relations management,”[78] and individual certification of acceptable data uses.[79] A concerted research effort to address the efficacy for either of these systems to achieve the goal of asserting personal data control in automated decision-making systems is sorely needed to answer this basic question of data ethics.

5.2 Implementable Data Ethics: Trust

The COVID-19 pandemic demonstrates that trust in institutions, such as the government, science, and technology firms, intersects with trust in persons in unexpected ways.[80] Identifying the factors which drive trust in science and data-driven decision-making is a notable post-pandemic challenge.[81] Improving the governance of technology, technology companies, and data flows between for-profit and research organizations represent overlapping areas in which trust could be studied and improved.[82] Other mechanisms include the creation of open data access points or windows into government activities and government-sponsored activities.[83]

Trust in technology is arguably linked to uses of privacy-enhancing technologies, encryption technologies, and companies’ methods for governing users’ permissions for companies’ data uses. Which variables drive trust, and in which settings? Do opt-in settings improve trustworthiness when the explicability of data used is otherwise low?[84] Do guarantees of consumer protection against risk from data breaches improve trust even when personal data gathered is sensitive?

[78] Searls, D. no date. “Vendor Relations Management” https://cyber.harvard.edu/people/dsearls

[79] Young, Kaliya and Yang, Lucy. 2020. "The COVID-19 Credentials Initiative: Bringing Emerging Privacy-Preserving Technology to a Public Health Crisis." Privacy & Pandemics Workshop Working Paper 10. Future of Privacy Forum, October 27.

[80] Dennis, Simon, et.al. (2020). “Social Licensing of Privacy-Encroaching Policies to Address the COVID-19 Pandemic”. Available at: https://stephanlewandowsky.github.io/UKsocialLicence/index.html

[81] Scarfuto, Jessica. February 16, 2020. “Do you Trust Science? These 5 Factors Play a Big Role” Available at: https://www.sciencemag.org/news/2020/02/do-you-trust-science-these-five-factors-play-big-role

[82] Simpson, Erin and Collins, Sara. 2020. "Trust Deficit Why a Lack of Trust in Government and Technology has Harmed our Pandemic Response." Privacy & Pandemics Workshop Working Paper 43. Future of Privacy Forum, October 27.

[83] Matthe Caramancion, Kevin. 2020. "The Disputed Inequality by Open Linked Data: A Position Paper." Privacy & Pandemics Workshop Working Paper 17. Future of Privacy Forum, October 27; Chin, Stephanie H. and Chin, Caitlin T. 2020. "Managing Future Pandemics: Benefits and Challenges of Creating a Common Data Space for Highly-Infectious Diseases." Privacy & Pandemics Workshop Working Paper 19. Future of Privacy Forum, October 27.

[84] Lodders, Adam and Paterson, Jeannie. 2020. "Opt-in Didn’t Work for Digital Contact Tracing in Australia, but does that Justify an ‘Opt-out’ Approach? Not yet – at Least in Australia." Privacy & Pandemics Workshop Working Paper 24. Future of Privacy Forum, October 27.

5.3 Implementable Data Ethics: Scope

The combination of the COVID-19 pandemic and the associated infodemic exposed the cross-fertilizations and complex interdependencies of contemporary data flows. The potential degree of transformation of health data through the associated “infodemic” highlights gaps in the implementation of ethical principles for the many actors invested in the production and use of COVID-19 research. For example, when ethically and methodologically sound research was misappropriated by nefarious actors for dubious purposes, questions emerged about the locus of responsibility for the flow of that research information. Identifying the patterns of research misappropriation in a crisis of this type or scale is a research challenge that could be met through work by bibliometrics, communications, and computer science researchers.

The speed of research efforts strained systems normally present to ensure the soundness of research. Preprint archives, self-publication through blogs and aggregators, and rapid dissemination of research all served to put large amounts of information into public hands. Yet the lack of oversight and controls led to some premature publication of conjectures or conclusions that did not withstand further scrutiny. Alas, such issues may jeopardize public trust and efforts to use research to inform policy. Identifying new publication ethics for research in a crisis is a necessary addition to discussions of publication norms and research ethics.

5.4 Implementable Data Ethics: Equity

Data analytics are powerful tools that have a significant effect on individuals’ lives. During times of crisis, the power shifts to those with access to data and analytics tools can surface, sometimes in alarming ways, situations of inequity that were present before the crisis erupted. Balancing power and equitably distributing risks of harms during a crisis situation requires careful consideration and design of “socio-legal-technical data sharing and management systems that accept heterogeneous, sensitive, and potentially biased data from both public and private sources as input; facilitate manual and computational procedures for data transformation, cleaning, linking, and publishing; produce integrated (and bias-adjusted) data products (e.g., visualizations, integrated datasets, trained models) as output; and manage access to the data, data products, and provenance to protect privacy, facilitate accountability, and generally enforce compliance with relevant law.”[85] The effects of such a socio-legal-technical system can be measured by assessing equity as a multidimensional concept amenable to definition mathematically, computationally, and experimentally:[86] representational equity, feature equity, access equity, and outcome equity.

[85] Howe, Bill, Jagadish, H. V., and Stoyanovich, Julia. 2020. "The Future of Privacy is Data Equity." Privacy & Pandemics Workshop Working Paper 11. Future of Privacy Forum, October 27, p. 1.

[86] Mellers, Barbara A. "Equity judgment: A revision of Aristotelian views." *Journal of Experimental Psychology: General* 111, no. 2 (1982): 242; Mandell, Marvin B. "Modelling effectiveness-equity trade-offs in public service delivery systems." *Management Science* 37, no. 4 (1991): 467-482; Howe, Bill, Jagadish, H. V., and Stoyanovich, Julia. 2020. "The Future of Privacy is Data Equity." Privacy & Pandemics Workshop Working Paper 11. Future of Privacy Forum, October 27, p. 1.

Equity pertains to more than just individuals when compared to other individuals. Equity between groups is also important to ensure that no communities disproportionately suffer risks due to the crisis itself or to the uses of their data in crisis response. Addressing group-based equity invites novel forms of risk identification, including privacy risk. Estimates of collective risks and collective harms, whether in the form of breaches of group privacy or mutual privacy, or differential impacts of representation in open datasets for training algorithms, remain an open area for research at the intersection of privacy, ethics, and data analysis.[87]

5.5 Implementable Data Ethics: Integrity

Data integrity is a concept with multiple meanings in the field of applied ethics. In the situation of research, it means that data is complete, verified, undistorted, and holds up to the scrutiny of peers.[88] In the situation of sensors and data analytics, such as might be found in an electric grid, data integrity means that a sensor system is free of manipulation or control by malicious actors.[89] Public health surveillance technology has transformed the public health research environment from one of human production of data to sensor production of data. This transformation suggests that the meaning of data integrity in public health surveillance will need to evolve; instead of examining strictly human faults in the data system, sensor level faults will need to be included. A system-level understanding of public health sensors and data technology is essential to learn when the system is operating at an optimal rate of disease detection and mitigation, when manipulated data inputs compromise the health, safety, and wellbeing of the population, and when the data being produced by sensors under attack must be modeled in alternative ways to yield actionable and accurate predictive insights.[90] The changing characteristics of a sensor and technology-enabled public health surveillance system means that a new focus of research must be introduced to measure the "state" of this new system accurately.

5.6 Implementable Data Ethics: Proportionality

Early in the pandemic, scholars asked the community of public health professionals, technology companies, and other researchers to imbue into their practices the clear and careful use of data to address the COVID-19 pandemic. While the crisis situation invited loosened restrictions of uses of data, whether increasing data access and thus risk to persons' personal data, led to benefits for persons and communities, is still an open question.[91] The growing concerns for proportionality, including accurately estimating whether a data or technologically-driven solution is "justified,"

[87] Puri, Anuj. 2020. "Post-Pandemic Simveillance." Privacy & Pandemics Workshop Working Paper 41. Future of Privacy Forum, October 27.

[88] National Academy of Sciences (US), National Academy of Engineering (US) and Institute of Medicine (US) Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington (DC): National Academies Press (US); 2009. 2, Ensuring the Integrity of Research Data. Available from: https://www.ncbi.nlm.nih.gov/books/NBK215260/

[89] S. Sridhar and G. Manimaran, "Data integrity attacks and their impacts on SCADA control systems," *IEEE PES General Meeting*, Providence, RI, 2010, pp. 1-6, doi: 10.1109/PES.2010.5590115.

[90] Luo, Jian, Tao Hong, and Shu-Cherng Fang. "Benchmarking robustness of load forecasting models under data integrity attacks." *International Journal of Forecasting* 34, no. 1 (2018): 89-104.

[91] Hoffman, David. Verbal Presentation. Privacy & Pandemics. October 28, 2020.

particularly when there is an inequitable distribution of outcomes, is the signal concern for data ethics and the study of science and technology in a society in crisis.[92]

6.0 Privacy-Preserving Public Health Surveillance Technologies

Public health surveillance is the "ongoing, systematic collection, analysis and interpretation of health-related data essential to planning implementation and evaluation of public health practice."[93] Public health surveillance requires the collection of sensitive or personal data to glean insights into the health-related status, behaviors, and outcomes of individuals and their families, partners, and communities. Measures taken to collect, use, secure, and share sensitive or personal data used in public health research and public health activities vary widely. This variation is accompanied by varying privacy protections for public health surveillance data and technologies.

6.1. Public Health Surveillance Data

The COVID-19 pandemic demonstrates the extent to which numerous streams of data can be drawn upon to advance public health initiatives. For example, common consumer technologies, such as Bluetooth in smartphones and Bluetooth beacons in retail settings, can be used to estimate proximity for exposure notification apps.[94] Commercial datasets, such as consumer credit card and loyalty program data, have been used to monitor movements and exposures.[95] Even data drawn from public health and public service infrastructures, such as sewage and transit system airflow, were drawn upon to inform public health initiatives and public policy decision-making.[96]

The pandemic also highlights challenges to building and relying on real-time, longitudinal public health surveillance data. Longitudinal data is especially important as it informs the development of novel computational solutions that estimate broader societal impacts, including the "aftershocks" of a crisis. To be most useful, though, longitudinal data must be complete and consistent across time and location. Ensuring consistency of data reporting across all governmental jurisdictions, hospitals, and health systems is a well-documented challenge that, if

[92] Matthe Caramancion, Kevin. 2020. "The Disputed Inequality by Open Linked Data: A Position Paper." Privacy & Pandemics Workshop Working Paper 17. Future of Privacy Forum, October 27.

[93] CDC "Public Health 101 Series, Introduction to Public Health Surveillance"

[94] Bater, Johes, et al. 2020. "Poirot: Private Contact Summary Aggregation." Privacy & Pandemics Workshop Working Paper 27. Future of Privacy Forum, October 27; Hernández-Orallo, Enrique, Carlos T. Calafate, Juan-Carlos Cano, and Pietro Manzoni. "Evaluating the Effectiveness of COVID-19 Bluetooth-Based Smartphone Contact Tracing Applications." *Applied Sciences* 10, no. 20 (2020): 7113.

[95] Chetty, Raj, John N. Friedman, Nathaniel Hendren, and Michael Stepner. "Real-time economics: A new platform to track the impacts of COVID-19 on people, businesses, and communities using private sector data." *NBER Working Paper* 27431 (2020).

[96] Barcelo, Damia. "An environmental and health perspective for COVID-19 outbreak: Meteorology and air quality influence, sewage epidemiology indicator, hospitals disinfection, drug therapies and recommendations." *Journal of Environmental Chemical Engineering* (2020): 104006; Peccia, Jordan, Alessandro Zulli, Doug E. Brackney, Nathan D. Grubaugh, Edward H. Kaplan, Arnau Casanovas-Massana, Albert I. Ko et al. "SARS-CoV-2 RNA concentrations in primary municipal sewage sludge as a leading indicator of COVID-19 outbreak dynamics." *medRxiv* (2020).

not corrected, can result in adverse actions, poor benchmarking of effective solutions, or failure to identify and address appropriately needs across diverse communities.

6.2 *Public Health Surveillance and Technology*

Data acquisition for public health surveillance is no longer an activity reserved for extensively trained and specialized public health professionals. Data from personal consumer technology, such as smartphones, fitness wearables, and commercially available scanners (e.g., thermal scanners), has been repurposed to support real-time or real-world public health surveillance and broad public safety uses.[97] These new types of public health data include information from "exposure notification apps," also called "digital contact tracing apps," produced in response to the COVID-19 pandemic. These apps gave billions of individuals the opportunity to engage in real-time or real-world public health surveillance personally.[98] To guide technological initiatives, the Centers for Disease Control and Prevention (CDC) laid out guidance in the form of technical questions for researchers to answer, such as how to effectively streamline data collection and management, particularly for laboratory testing activities, in ways that do not disrupt or divert resources.[99] The CDC also pressed for innovations to improve data acquisition, cleaning, and uses of data for real-time monitoring of the performance of surveillance activities at scale.

6.2.1 Control and Efficacy of New Technologies

From the perspective of data privacy and security, the differences in methods and degrees of user control over the technology-centered acquisition of public health surveillance data confounds efforts to establish clear and consistent pathways for personal control of information.[100] Ideally, policy guidance would outline technical guidance in this area. For example, the transmission of personal health and public health-related information between devices and/or centralized devices and cloud servers opens opportunities for malicious, curious, or even research-driven intrusion. Further, clear guidance would outline limitations necessary to limit or harness crosstalk and information leaks between devices for positive purposes. Thus, the security of both on-device and centralized personal health and health-related information is an area that is ripe for convergent research.[101]

[97] Hamilton, Janet J. and Hopkins, Richard S. (2018). *The CDC Field Epidemiology Manual.* Available at: https://www.cdc.gov/eis/field-epi-manual/chapters/data-collection-management.html

[98] Altshuler, Tehilla Shwartz and Aridor-Hershkovitz, Rachel. 2020. "Evaluating the Efficiency of Contact Tracing Technologies." Privacy & Pandemics Workshop Working Paper 9. Future of Privacy Forum, October 27; Díaz Díaz, Efrén. 2020. "Geolocation Apps do not Cure COVID-19: They Analyse People's Mobility (Case of Spain)." Privacy & Pandemics Workshop Working Paper 26. Future of Privacy Forum, October 27; Pierson, Jo. 2020. "Contact Tracing Apps and Solutionism." Privacy & Pandemics Workshop Working Paper 7. Future of Privacy Forum, October 27;

[99] Hamilton, Janet J. and Hopkins, Richard S. (2018). *The CDC Field Epidemiology Manual.* Available at: https://www.cdc.gov/eis/field-epi-manual/chapters/data-collection-management.html

[100] Palfrey, Quentin, Ghamrawi, Lena, Good, Nathan, and Monge, Will. 2020. "COVID-19 Mobile App Accountability Investigation Recommendations." Privacy & Pandemics Workshop Working Paper 33. Future of Privacy Forum, October 27.

[101] Agbinya, Johnson I. "Analysis of New Crosstalk Multiple Access Systems for Inductive Communication Systems". *Proceedings of the 2nd Pan African International Conference on Science, Computing and Telecommunications (PACT 2014),* Arusha, 2014, pp. 17-22, doi: 10.1109/SCAT.2014.7055124.

Did any of the methods used to preserve the privacy of public health surveillance data work as expected? Which were most efficacious, and which factors drove their comparative efficacy? Measuring the efficacy of pandemic control and containment techniques, estimating the duration and extent of harms to individuals and groups that emerged due to these technologies, fostering better transparency for private and government entities use of these technologies, and establishing which techniques could be effectively deployed for privacy in future crises, are important. Answering these questions was previously the task of healthcare organizations such as the National Institutes of Health. However, the strong influence of computational sciences and technology during the COVID-19 pandemic means that estimation of efficacy and harm is also a challenge for data-driven disciplines. For example, estimating the number of intrusions into research databases, estimating the privacy loss for individuals who used pandemic apps, and estimating the costs and benefits to computer science research team of attempts to secure privacy in COVID-19 research data effectively could lead to refinements in the approach to use of computing resources during future crises.[102]

6.2.2 Researching Novel Public Health Surveillance Technologies

Contact tracing apps received most of the public's attention for COVID-19 technologies. In the meantime, researchers building fully integrated Wireless Sensor Networks (WSNs), while coupling their research with enhanced investigations that securitize sensor devices and the Internet of Things (IoT), presented a research challenge for the future of health surveillance.[103] WSNs, including body area networks (BANs) or "a network consisting of intelligent, low power, micro and nano-technology sensors and actuators, which can be placed on the body or implanted in the human body, providing timely data," represent an unexplored area of public health surveillance technology as a resource for mitigation of COVID-19.[104] Although consumer wellness wearables were explored as a source of biometric data for detection of SARS-CoV-2 infection, the promise of a fully-integrated system of WSNs across the multiple domains did not come to fruition. Multi-domain research is necessary to ensure that the implementation of such technologies and data can inform not just public health surveillance but also the development and sustainability of other important social safety nets, like transportation infrastructure and precision agriculture programs, to mitigate the myriad economic challenges imposed by COVID-19.

7.0 Conclusions

The public debate over the creation and uses of exposure notification apps has highlighted the importance of ethics, including privacy, at all technological development stages. Beyond discussions of contact tracing as such, the COVID-19 pandemic has elevated the importance of many other components of responsible uses of technology, including management of

[102] K. Yelick. Keynote Presentation. Privacy & Pandemics. October 28, 2020. Virtual Conference.

[103] Llamas, Lyz. 2020. "IoMT and Data Protection Challenges during the COVID-19 Pandemic." Privacy & Pandemics Workshop Working Paper 38. Future of Privacy Forum, October 27.

[104] Movassaghi, Samaneh, Mehran Abolhasan, Justin Lipman, David Smith, and Abbas Jamalipour. "Wireless body area networks: A survey." *IEEE Communications surveys & tutorials* 16, no. 3 (2014): 1658-1686, p. 1658-1659.

expectations, public and private ownership, government uses and government support, and public-private partnerships and collaboration.[105] The pandemic has also highlighted how public and private actors favor technological solutions over political and economic solutions.[106] Technological solutionism presses concerns about effective control of technology companies by political and economic authorities and the citizens who (in some places) elect said authorities.

The technologies that have evolved to take on new roles in our lives or emerged as a consequence of the global pandemic are not "neutral" feats of engineering that influence our movements and desires infrequently, and are remote for the majority of individuals. Pandemic technologies are socio-technical systems designed to be an embedded part of the experience of social life, education, healthcare, and work. They are built to adapt to changes to these social systems by taking in data in myriad forms and turning it into usable information for individuals, governments, schools, and workplaces to use when making decisions on further adaptation and response to COVID-19.[107]

The global pandemic highlights clearly how enormous the gauge of the needle of data is that sticks into the arm of our lives. As the public grappled with this realization, calls for responsible uses of data grew louder. Responsible use of data is no longer an academic consideration, it is an immensely personal and practical one. The next round of Convergence Accelerator ("C-Accel") proposals can bring research to bear on this most pressing and practical of problems.

FPF has proposed Privacy & Pandemics: Responsible Use of Data During Times of Crisis as the focus of NSF's latest C-Accel program.[108] Our C-Accel proposal is driven by an urgent need to develop effective structures, protocols, and processes for data sharing and governance while protecting individual privacy in a post-pandemic world. Given NSF's leadership on data science research and development, and role in advancing US leadership in health and science around the world, NSF can lead the way in preparedness for future pandemics by using its resources to improve the quality of data necessary to power healthcare, public health decision-making, and public policy choices.

Convergence power and research collaboration of academics, industry practitioners, and public officials can fundamentally transform how data is used, managed, and protected in support of public health. Specific endeavors would include establishing convergence teams to develop

[105] The Future of Privacy Forum's Testimony before the Senate Committee on Commerce, Science, and Transportation, "Enlisting Big Data in the Fight Against Coronavirus" (Apr. 9, 2020), https://www.commerce.senate.gov/services/files/F24D0AF8-D939-4D14-A963-372B9357DD7E.4D14-A963-372B9357DD7E.

[106] Pierson, Jo. 2020. "Contact Tracing Apps and Solutionism." Privacy & Pandemics Workshop Working Paper 7. Future of Privacy Forum, October 27.

[107] Hawn Nelson, Amy, et al. 2020. "Position Statement." Privacy & Pandemics Workshop Working Paper 18. Future of Privacy Forum, October 27.

[108] *See* FPF, "Privacy & Pandemics: Responsible Use of Data During Times of Crisis" (May 29, 2020) (FPF's response to C-Accel RFI).

best-practices guidance and data-based policy formulation for responsible data use in support of health initiatives; encouraging interoperability of the many data sources that inform data-driven healthcare decision-making; creating FAIR data repositories; facilitating multi-institutional sharing of related data science education and research expertise; and accelerating programs for ethical sharing of data across industry-academic collaboratives.

We[109] conclude with two important notes about sustainability:

1. The proposed track will leverage and expand NSF's other COVID-19-related investments toward a goal of improved data sharing and governance with individual privacy rights and protections at the forefront.
2. The Privacy and Pandemics conference illustrated the value of shared forums like this to define and marry technology needs with good policy/policy frameworks. Both are needed to accomplish system-level changes and achieve long-term sustainability.

[109] Composition of this report was aided by: Dr. Gabriela Zanfir-Fortuna, Dr. Rachele Hendricks-Sturrup, Ms. Barbara Kelly, Ms. Brenda Leong, Ms. Caroline Hopland, Mr. Jeremy Greenberg, Ms. Kelsey Finch, and Ms. Veronica Alix.

Privacy and Pandemics Workshop Participants

https://fpf.org/2020-pandemics-conference

October 27, 2020 Day 1- Technology & Data Access

Welcome & Keynote

10:00am - 10:40am

Name	**Affiliation**
Hosts:2	
Christy Harris	Director of Technology & Privacy Research, Future of Privacy Forum
Jules Polonetsky	CEO, Future of Privacy Forum
Conveners: 2	
David Hoffman	Professor of the Practice, Duke Sanford School of Public Policy, & Associate General Counsel and Senior Director of Data Policy, Intel Corporation
Chaitan Baru	PhD., Senior Science Advisor, National Science Foundation
Keynote:	
Dr. Lauren Gardner	Associate Professor of Civil Engineering, CSSE Co-Director, Johns Hopkins University

Session 1: Accessibility of SARS CoV-2 Data

10:40am - 12:10pm

The need to understand, research, and prevent the spread of COVID-19 has necessitated urgent efforts to assemble, collect, manage, and transfer volumes of data from a variety of disparate sources. These efforts have raised unique challenges related to data access and quality, systems' interoperability, and display & visualization of information. In this session, experts will grapple with the ongoing challenges related to data access and the corresponding privacy concerns.

Name	**Affiliation**
Moderator:	
Kelsey Finch	Senior Counsel, Future of Privacy Forum
Firestarters:5	**Affiliation**
Bill Howe	Associate Professor in the Information School at the University of Washington, and Director of the Urbanalytics Group
Nicole Contaxis	Data Librarian and Lead of Data Discovery at NYU Langone Health
Anthony (Tony) Reina	Intel's Chief AI Architect for Health & Life Sciences, Intel Corporation
Khaled El Emam	CHEO Research Institute and the University of Ottawa
Aaron Katz	Johns Hopkins Applied Physics Laboratory (APL)

Participants: 13	
Jamie Blackport	Founder & CEO, Mirador Analytics
Kevin Matthe Caramancion	College of Emergency Preparedness, Homeland Security, and Cybersecurity, University at Albany, State University of New York
Stephanie Chin	Ph.D. Candidate, Stanford University, Department of Civil and Environmental Engineering
Sara Collins	Policy Counsel, Center for American Progress & Public Knowledge
Dov Greenbaum	Director, Institute for Legal Implications of Emerging Technologies @ the Interdisciplinary Center Herzliya (IDC) in Israel & the Gerstein Laboratory @ Yale University
Abhishek Gupta	Founder & Principal Researcher, Montreal AI Ethics Institute, Machine Learning Engineer & CSE Responsible AI Board Member, Microsoft
Sinead Impey	ADAPT Centre, School of Computer Science and Statistics, Trinity College, Dublin
Lyz Llamas	MLB, Senior Privacy Counsel, FIRST PRIVACY GmbH
Niamh McGuinness	Senior Analyst on the Clinical Trial Transparency Team at Privacy Analytics
Kathryne Metcalf	Graduate Fellow, Communication and Science Studies University of California San Diego
Ryan Naughton	Executive Director of the COVID Alliance
Professor Srinivasan Parthasarathy	Computer Science and Engineering, Director, Data Mining Research Laboratory, Co-Director, Data Analytics Program, Ohio State University
Kaliya Young	Identity Woman, Co-Founder of the Internet Identity Workshop, Merritt College, Covid Credentials Initiative

Session 2: Use of Technology to Track, Trace, and Notify to Control the Spread of COVID-19
12:30pm - 2:00pm

Identifying the mechanisms for spread of COVID-19 presents scientific and social challenges. Concerns about transmission through direct contact, shared surfaces, and via airborne droplets led to the development of new technologies to facilitate appropriate social distancing and isolation following an infection, contact tracing and exposure notification apps, thermal scanning, and self-isolation symptom recommendation services to reduce person-to-person exposure. These technologies raise concerns about data protection, privacy, and public trust, which experts will address in this panel.

Name	**Affiliation**
Moderator:	
Polly Sanderson	Policy Counsel, Future of Privacy Forum
Firestarters:4	
Jeannie Paterson	Co-Director of the Centre for AI and Digital Ethics, University of Melbourne
Anne L. Washington	PhD, Assistant Professor of Data Policy, Applied Statistics, Social Science, and Humanities, Steinhardt School, New York University
Chas Kissick	Duke University Sanford School of Public Policy
Dr. Edoardo Celeste	Assistant Professor in Law, Technology and Innovation at the School of Law and Government of Dublin City University

Participants:13	
Sam Andrey	Director of Policy & Research, Ryerson Leadership Lab, Toronto, Canada
Rachel Aridor-Hershkovitz	The Center for Democratic Values and Institutions, Researcher |The Israel Democracy Institute
Dr Garfield Benjamin	Postdoctoral Researcher, Solent University, UK
Efrén Díaz Díaz	Lawyer, Geospatial Law PhD. Responsible for the Areas of Technology and Space Law, Bufete Mas y Calvet
Phoebe Dijour	Duke University Bass Connections, The Privacy Implications of COVID-19 Contact Tracing Technology Team
Jiyeon Kim	Research Fellow, Cordell Institute, Washington University School of Law
Kartik Nayak	Assistant Professor in Computer Science, Duke University
Quentin Palfrey	International Digital Accountability Council
Prof. Jo Pierson	Holder of the Chair on Data Protection On The Ground at the Vrije Universiteit Brussel, Belgium
Joy Pritts	2020-2021 Innovators Network Foundation Privacy Fellow (Former Chief Privacy Officer at the Office of the National Coordinator for Health Information Technology within the U.S. Department of Health and Human Services)
Brian Ray	Leon M. and Gloria Plevin Professor of Law, Director, Center for Cybersecurity and Privacy Protection, Cleveland-Marshall College of Law, Cybersecurity Liaison, IoT Collaborative
Virág Réti	Chief Executive Officer, Xtendr
Laurin B. Weissinger	Fletcher School, Tufts University, Cyber Security Fellow, Yale Law School, Visiting Fellow, Information Society Project, Yale Law School

Day 2: October 28, 2020 - Policy & Convergence

Welcome & Keynote

10:00am - 10:40am

Hosts:	**Affiliation**
Christy Harris	Director of Technology & Privacy Research, Future of Privacy Forum
Jules Polonetsky	CEO, Future of Privacy Forum
Convener:	
Dr. Rob Brennan	Assistant Professor, School of Computing, Dublin City University
Keynote:	
Dr. Kathy Yelick	Associate Dean for Research, Division of Computing, University of California, Berkeley

Session 3: Adapting Legal and Regulatory Responses to a Global Emergency
10:40am - 12:10pm

The contention that law and regulation struggles to keep pace with technology have become ever more prominent as the COVID-19 pandemic emergency grows. Privacy laws, public health authority provisions, and international law all emerged as areas of concern. In this session, experts discuss what we can learn from this pandemic for the future of law, regulatory authority, and social norms.

Name	**Affiliation**
Moderator:	
Limor Shmerling Magazanik	Managing Director, Israel Tech Policy Institute
Firestarters:3	
Evan Selinger	Professor of Philosophy, Rochester Institute of Technology
Kimball Dean Parker	CEO at SixFifty and Director of LawX at Brigham Young University
Magali (Maggie) Feys	AContrario.Law - Anonos
Participants:13	
Haleh Asgarinia	PhD Candidate, University of Twente | Faculty Behavioral, Management and Social Science | Department of Philosophy
Jill Bronfman	Privacy Counsel, Common Sense Media
Oskar van Deventer	Senior Scientist, TNO
Beverley Hatcher-Mbu	Senior Associate, Development Gateway, Inc.
Rachele Hendricks-Sturrup	Health Policy Counsel, Future of Privacy Forum
Deanne Kasim	Executive Director, Health Policy
Tiffany C. Li	Fellow at Yale Law School's Information Society Project, Visiting Clinical Assistant Professor at Boston University School of Law
Nora McDonald	Research Assistant Professor, University of Maryland, Baltimore County
Amy Hawn Nelson	Director of Training and Technical Assistance, Actionable Intelligence for Social Policy, University of Pennsylvania
Anuj Puri	PhD student, St Andrews and Stirling Graduate Programme in Philosophy (SASP), University of St Andrews
Divya Ramjee	PhD Candidate & Adjunct Professor, Department of Justice, Law and Criminology, School of Public Affairs, American University
Boris Segalis	Partner, Cooley LLP, Vice Chair cyber/data/privacy practice
Bhumika Sharma	Ph.D. Research Scholar, Himachal Pradesh University, Shimla, India
Omer Tene	Chief Knowledge Officer, International Association of Privacy Professionals

Session 4: The Convergence of Technology, Policy, and Responsible Data Use in a Global Crisis

12:30pm - 2:00pm

The COVID-19 pandemic brought issues of privacy and technology to the forefront of the public and political debate. The lessons from this pandemic will seed future research, but in which directions ought privacy law, technology, and research go when interfacing with the many scientific communities that will explore these questions? In this final workshop, the Future of Privacy Forum invites discussion and debate around our position statement on the future of privacy technology and research.

Name	**Affiliation**
Moderator:	
Jules Polonetsky	CEO, Future of Privacy Forum
Discussants:4	
Dr. Sara Jordan	Policy Counsel, Artificial Intelligence and Ethics, Future of Privacy Forum
Vincent Cassard	International Committee for the Red Cross
Natalie Evans Harris	Head of Strategic Initiatives, Brighthive
David Hoffman	Intel Corp. & Duke Sanford School of Public Policy
Panelists: 46	
Sam Andrey	Director of Policy & Research, Ryerson Leadership Lab, Toronto, Canada
Rachel Aridor-Hershkovitz	The Center for Democratic Values and Institutions, Researcher |The Israel Democracy Institute
Haleh Asgarinia	PhD Candidate, University of Twente | Faculty Behavioral, Management and Social Science | Department of Philosophy
Jane Bambauer	Professor of Law, University of Arizona
Dr. Malika Bedechache	Assistant Professor in the School of Computing at Dublin City University
Dr Garfield Benjamin	Postdoctoral Researcher, Solent University, UK
Jamie Blackport	Founder & CEO, Mirador Analytics
Jill Bronfman	Privacy Counsel, Common Sense Media
Kevin Matthe Caramancion	College of Emergency Preparedness, Homeland Security, and Cybersecurity, University at Albany, State University of New York
Caitlin T. Chin	Research Analyst, The Brookings Institution, Center for Technology Innovation
Ana DeCesare	Duke University Bass Connections, The Privacy Implications of COVID-19 Contact Tracing Technology Team
Oskar van Deventer	TNO
Efrén Díaz Díaz	Lawyer, Geospatial Law PhD. Responsible for the Areas of Technology and Space Law, Bufete Mas y Calvet
Khaled El Emam	CHEO Research Institute and the University of Ottawa

Jeremy Epstein	Lead Program Officer, Secure and Trustworthy Cyberspace, National Science Foundation
Magali (Maggie) Feys	AContrario.Law - Anonos
Abhishek Gupta	Founder & Principal Researcher, Montreal AI Ethics Institute, Machine Learning Engineer & CSE Responsible AI Board Member, Microsoft
Beverley Hatcher-Mbu	Senior Associate, Development Gateway, Inc.
Amy Hawn Nelson	Director of Training and Technical Assistance, Actionable Intelligence for Social Policy, University of Pennsylvania
Bill Howe	Associate Professor in the Information School at the University of Washington, and Director of the Urbanalytics Group
Chris Keaton	OneTrust
Jiyeon Kim	Research Fellow, Cordell Institute, Washington University School of Law
Chas Kissick	Duke University Sanford School of Public Policy
Tiffany C. Li	Fellow at Yale Law School's Information Society Project, Visiting Clinical Assistant Professor at Boston University School of Law
Lyz Llamas	MLB, Senior Privacy Counsel, FIRST PRIVACY GmbH
Adam Lodders	Academic and Research Programs Manager, Centre for AI and Digital Ethics, University of Melbourne
Nora McDonald	Research Assistant Professor, University of Maryland, Baltimore County
Niamh McGuinness	Privacy Analytics
Kathryne Metcalf	Graduate Fellow, Communication and Science Studies University of California San Diego
Ryan Naughton	Executive Director of the COVID Alliance
Kartik Nayak	Assistant Professor in Computer Science, Duke University
Kimball Dean Parker	CEO at SixFifty and Director of Law X at Brigham Young University
Prof. Srinivasan Parthasarathy	Computer Science and Engineering, Director, Data Mining Research Laboratory, Co-Director, Data Analytics Program, Ohio State University
Joy Pritts	2020-2021 Innovators Network Foundation Privacy Fellow (Former Chief Privacy Officer at the Office of the National Coordinator for Health Information Technology within the U.S. Department of Health and Human Services)
Anuj Puri	PhD student, St Andrews and Stirling Graduate Programme in Philosophy (SASP), University of St Andrews
Divya Ramjee	PhD Candidate & Adjunct Professor, Department of Justice, Law and Criminology, School of Public Affairs, American University
Mario Romao	Director, EU AI and Global Healthcare Policy, Intel Corporation
Evan Selinger	Professor of Philosophy, Rochester Institute of Technology
Erin Simpson	Associate Director, Technology Policy at Center for American Progress
Luke Stark	Assistant Professor, Faculty of Information & Media Studies (FIMS), University of Western Ontario
Julia Stoyanovich	Assistant Professor
Ferenc Vágujhelyi	Xtendr Zrt.

Ine Van Zeeland	PHD Researcher at imec-SMIT, Vrije Universiteit Brussel
Dr. P.J. Wall	Research Fellow with the INTEGRITY H2020 Project in the ADAPT Centre (www.adaptcentre.ie) in the School of Computer Science, Trinity College Dublin
Anne L. Washington	PhD, Assistant Professor of Data Policy, Applied Statistics, Social Science, and Humanities, Steinhardt School, New York University
Laurin B. Weissinger	Fletcher School, Tufts University, Cyber Security Fellow, Yale Law School, Visiting Fellow, Information Society Project, Yale Law School
Kaliya Young	Identity Woman, Co-Founder of the Internet Identity Workshop, Merritt College, Covid Credentials Initiative