# Importance of Participant-Centricity and Trust for a Sustainable Medical Information Commons

*Amy L. McGuire, Mary A. Majumder, Angela G. Villanueva, Jessica Bardill, Juli M. Bollinger, Eric Boerwinkle, Tania Bubela, Patricia A. Deverka, Barbara J. Evans, Nanibaa' A. Garrison, David Glazer, Melissa M. Goldstein, Henry T. Greely, Scott D. Kahn, Bartha M. Knoppers, Barbara A. Koenig, J. Mark Lambright, John E. Mattison, Christopher O'Donnell, Arti K. Rai, Laura L. Rodriguez, Tania Simoncelli, Sharon F. Terry, Adrian M. Thorogood, Michael S. Watson, John T. Wilbanks, and Robert Cook-Deegan*

Comprehending complex health diseases and conditions calls for the ability of big data analytics to analyze vast amounts of data on diverse health-related variables. Such data are typically sourced from multiple projects and data repositories, forming a data resource commons. The importance of data commons have not gone unnoticed.[1] A 2011 National Academies committee report called for the development of an Information Commons and Knowledge Network to advance research and improve health.[2] Over the past decade, we have seen the evolution of medical information commons (MIC) in the U.S. and elsewhere, which we define as networked environments in which diverse sources of data on large populations become broadly available for research use and clinical applications, and which include the collection of many different common pool resources.[3] As Elinor Ostrom observed over a career of studying various kinds of common pool resources, the stability and effectiveness of any commons requires forming a community, establishing rules, and monitoring compliance with those rules.[4] Some groups, such as the Global Alliance for Genomics and Health (GA4GH) have worked to establish rules to govern how the resources in an MIC are brought together. Informed

**Amy L. McGuire, J.D., Ph.D.**, *is the Leon Jaworski Professor of Biomedical Ethics and Director of the Center for Medical Ethics and Health Policy at Baylor College of Medicine. Dr. McGuire serves on the program committee for the Greenwall Foundation Faculty Scholars Program in Bioethics and is immediate past president of the Association of Bioethics Program Directors.* **Mary A. Majumder, J.D., Ph.D.**, *is an Associate Professor of Medicine at the Center for Medical Ethics and Health Policy, Baylor College of Medicine.* **Angela G. Villanueva, M.P.H.**, *is a Research Associate at the Center for Medical Ethics and Health Policy at Baylor College of Medicine.* **Jessica Bardill, Ph.D.**, *is an Assistant Professor of Indigenous Canadian Literatures and Cultures at Concordia University in Montreal, Quebec.* **Juli M. Bollinger, M.S.**, *is a Research Associate in the Center for Medical Ethics and Health Policy at the Baylor College of Medicine and a Research Associate and Associate Faculty at the Berman Institute of Bioethics at Johns Hopkins University.* **Eric Boerwinkle, Ph.D.**, *is the Dean of the University of Texas Health Science Center at Houston, School of Public Health and the Associate Director, Human Genome Sequencing Center at Baylor College of Medicine.* **Tania Bubela, Ph.D., J.D.**, *is the Dean of the Faculty of Health Sciences at Simon Fraser University and a Fellow of the Canadian Academy of Health Sciences.* **Patricia Deverka, M.D., M.S., M.B.E.**, *is Director, Value Evidence and Outcomes at Geisinger National Precision Health, where she focuses Outcomes at Geisinger National Precision Health, where she focuses systems and policymakers.* **Barbara Evans, MS, Ph.D., J.D., LL.M.**, *is the Mary Ann and Lawrence E. Faust Professor of Law and Director of the Center for Biotechnology & Law at the University of Houston Law Center and holds a joint appointment as Professor of Electrical and Com-*

by that work, and drawing on the research that we conducted over the past three years, which included a landscape analysis, expert stakeholder interviews, and community advisory panels in three cities across the U.S., we believe that the two most important features of an MIC, around which those rules must be established, are: (1) **an MIC must be oriented around the people whose data it contains** and whom an MIC is ultimately intended to benefit; and (2) **the system must be trustworthy**. Here we discuss these features and how to best address them in order to build a sustainable resource, or rather, a sustainable, useful and widely available collection of linked resources.

## What Does it Mean to be Participant-Centric?

A central determinant of the long-term effectiveness of an MIC will be its ability to meet the needs of the people it is supposed to serve. Most of the decisions that affect flow of information and materials are now made under models designed by and for academic research institutions, private health care delivery organizations, commercial laboratories and other institutions organized around generation and use of data. But do the interests of those who design the systems and hold the data align with the rights and interests of the participants — the people the data describe — who are also, ultimately, its intended beneficiaries? Throughout our work we heard a strong desire from both expert stakeholders and community members for an MIC to be participant-centric. In our interviews with expert stakeholders, for example, the overwhelming majority confirmed that it is important to give participants a significant role and that limited conceptions of the participant role in research are no longer tenable in the context of an MIC.[5] The overwhelming majority of individuals on our community advisory panels reached the same conclusion.[6] But what does it mean for an MIC to be participant-centric and how can it be accomplished?

Scholars describe participant-centric initiatives as "tools, programs, and projects that empower participants to engage in the research process" by giving participants the option to have control over data and to engage in a reciprocal partnership with researchers.[7] Fundamentally, being participant-centric means showing respect for participants *as* persons with a voluntary, continuous role in decision-making, versus human subjects who are only engaged during the consent process, or sets of data points without

*puter Engineering at the UH Cullen College of Engineering.* **Nanibaa' A. Garrison, Ph.D.** (**Navajo**), *is an Assistant Professor at the Treuman Katz Center for Pediatric Bioethics at Seattle Children's Hospital and Research Institute and at the Department of Pediatrics at the University of Washington School of Medicine.* **David Glazer** *is an engineering director at Verily Life Sciences, where he helps life science organizations use cloud computing to accelerate and scale their work with big data.* **Melissa M. Goldstein, J.D.**, *is an Associate Professor in the Department of Health Policy and Management at the Milken Institute School of Public Health at the George Washington University.* **Henry T. Greely, J.D.**, *is the Deane F. and Kate Edelman Johnson Professor of Law and a professor (by courtesy) of Genetics at Stanford University, where he directs the Center for Law and the Biosciences. He is President of the International Neuroethics Society, co-chair of the Neuroethics Work Group of the NIH BRAIN Initiative, and a member of the National Academies' Committee on Science, Technology, and Law.* **Scott D. Kahn, Ph.D.**, *is the Chief Information Officer at LunaDNA where he is responsible for all informatics and data science strategy.* **Bartha M. Knoppers, Ph.D.**, *(Comparative Medical Law), is a Full Professor, Canada Research Chair in Law and Medicine and Director of the Centre of Genomics and Policy of the Faculty of Medicine at McGill University.* **Barbara A. Koenig, Ph.D.**, *is Professor of Bioethics and Medical Anthropology, based at the Institute for Health & Aging, University of California, San Francisco. She also serve as Director of the UCSF Program in Bioethics.* **J. Mark Lambright, M.B.A.**, *is the Chairman, Trident Advisors, LLC - CEO (Interim), Private Access, Inc. Mark Lambright has been a senior executive of a number of healthcare and technology companies ranging in size from startup to multi-billion in revenue.* **John Mattison, M.D.**, *is Assistant Medical Director, KP, SCAL, Chief Health Information Officer. Founder of CDA/CCD (XML standard for healthcare interoperability). Co-Editor of Healthcare Information Technology (2017 McGraw Hill).* **Christopher J. O'Donnell, M.D., M.P.H.**, *is Chief, Cardiology Section and Director, Center for Population Genomics, of the Boston VA Healthcare System and co-Principal Investigator (Chief Scientist) of the national VA Million Veteran Program; and is VA Contact Principal Investigator of the NIH All of Us Research Program. He is also a faculty member of the Department of Medicine, Brigham and Women's Hospital and Associate Professor of Medicine, Harvard Medical School.* **Arti K. Rai, J.D.**, *is the Elvin R. Latty Professor of Law and co-Director of the Center for Innovation Policy at Duke Law.* **Laura L. Rodriguez, Ph.D.**, *is the Director of the Division of Policy, Communications, and Education at the National Human Genome Research Institute.* **Tania Simoncelli, M.S.**, *is the Director of Science Policy at the Chan Zuckerberg Initiative. Formerly Executive Director, Count Me In at the Broad Institute, Assistant Director of Forensic Science and Biomedical Innovation, White House Office of Science & Technology Policy, and Science Advisor to the American Civil Liberties Union.* **Sharon F. Terry, M.A.**, *is the president and CEO of Genetic Alliance.* **Adrian Thorogood, B.C.L./LL.B.**, *is a lawyer and Academic Associate at the Centre of Genomics and Policy at McGill University in Montreal, Canada. He manages the Regulatory and Ethics Work Stream of the Global Alliance for Genomics and Health.* **Michael S. Watson, M.S., Ph.D.**, *is the Executive Director, American College of Medical Genetics and Genomics and ACMG Foundation for Genetic and Genomic Medicine and an Adjunct Professor of Pediatrics at Washington University School of Medicine.* **John T. Wilbanks** *is the Chief Commons Officer at Sage Bionetworks, and co-PI on awards for AllofUs Research Program, ELSI Issues in Unregulated Mobile Research, CTSA Center for Data To Health.* **Robert Cook-Deegan, M.D.**, *is a Professor in the School for the Future of Innovation in Society at Arizona State University.*

interest or concern for how the data are used or analyzed. It is important to recognize that the principle of participant-centrism is not the result of armchair (or academic seminar) analysis or pie-in-the-sky thinking. In fact, existing citizen science genomics efforts such as the Personal Genome Project exemplify norm-disruptive approaches to engaging participants and data-sharing.[8] In addition, Elinor Ostrom and her colleagues' extensive empirical work establishes that a successful commons must involve key stakeholders in the governance of the shared resource.[9] For an MIC, participant-centricity begins with governance, ensuring that all key stakeholders — including the participants who are the data contributors and ultimate beneficiaries — have a seat at the table.

A participant-centric MIC also recognizes and mitigates the risks that participants, individually and collectively, face, such as re-identification, as well as discrimination and stigmatization based on the collection, aggregation, and use of genomic and other health-related data.[10] Risks of discrimination and stigmatization are of special concern to individuals and groups who have already experienced significant social disadvantage and are especially vulnerable to harm from impersonal, non-transparent algorithmic decision-making drawing on big data.[11] Concern may also be heightened for particular areas of genomic and other health-related information, such as information about potential genomic contributors to drug or substance abuse, propensity for criminal behavior, and intelligence or impulsivity, as well as particular kinds of research uses, such as studies that stratify by social or ancestry groups.[12] In the U.S., some risks could be reduced if the Genetic Information Nondiscrimination Act (GINA) were strengthened and expanded beyond health insurance and employment to include other forms of genetic discrimination in eligibility for life insurance or access resources.[13] Given these gaps and weaknesses in current health data privacy laws, obligations to respect the autonomy of participants and potential participants and to invest in initiative or project-level privacy protection measures are especially strong. For example, if MICs establish credible sanctions for unauthorized re-identification of data and violations of agreements about permissible data uses, that could help discourage careless or malicious acts that might result in harm to participants while offering those affected a means of redress.

Another aspect of being participant-centric is ensuring that participants' voices are heard and included in meaningful ways in the governance of their collective data so that they can ensure that there is alignment between uses and their values, needs, and interests. A "researcher reputation system" where participants can post ratings on researchers and a report card on researchers' dissemination of results to participants and open-access journals, along with other metrics, has been suggested to cultivate this type of engagement.[14] Other strategies to ensure that participants have a voice at the table should also be explored.

Ethically and pragmatically, then, for its long-term sustainability as a rich data resource, it is in the interest of an MIC to be participant-centric. Being participant-centric can build public support and improve recruitment and retention, inclusive of historically underrepresented groups.[15] It can also help ensure that priorities, practices and outcomes align with public and participants' expectations and values, in order to increase and sustain trust, as discussed in more detail below. At the same time, the degree to which participants are involved in existing data resources falls on a continuum. At the far end are fully participant-driven initiatives. For example, through organizations such as PXE International (https://www.pxe.org) and the Chordoma Foundation (https://www.chordomafoundation.org), affected individuals and families have systematically created online disease communities with a common purpose, generated resources needed to advance research and clinical care, and set the terms for provision of these resources to researchers and private firms while themselves conducting citizen science.[16] The Life Raft Group (https://liferaftgroup.org) has similarly created data and other research resources for gastro-intestinal stromal tumors.[17] Under the tagline, "Your data drives discovery" the Multiple Myeloma Research Foundation (https://themmrf.org) has adopted novel business practices derived from industry to establish a database, collected over 4,000 samples in a repository, refined diagnostic technologies, raised funds for research, sponsored sequencing projects, and cultivated industry partners.[18] This initiative has been transformational by increasing life expectancy, in part through the ten drugs that have made it to market and the many more drugs that are still being tested. The Cystic Fibrosis Foundation (https://www.cff.org) funded early work, provided data access to patients, and made resources available that led to the suite of drugs now available to treat most forms of the disease.[19] However, that model has faltered at the point of access to treatment due to pricing of the end-products, with salient controversies over coverage and reimbursement in the U.K., Australia, Ireland, France, Canada, and the U.S. To date, most of the strongest participant-driven projects focus on cancer or rare, often inherited, diseases.[20]

Many valuable lines of research do not map to a constituency that can be organized around a single disease, however. For example, people with genomic variants

usually associated with disease pathology, but who are actually healthy, may harbor clues about the underlying biology that cannot be discovered by only studying those with disease. Understanding factors that result in avoiding disease or determining age of onset will depend on data and materials from unaffected individuals, who are followed longitudinally. Large-scale studies can also reveal an individual's risk of various health outcomes, which when linked to therapeutic interventions should guide individualized wellness and prevention programs. This necessity is one of the compelling rationales for large-scale national genomic projects, such as All of Us (https://allofus.nih.gov) and the UK Biobank (https://www.ukbiobank.ac.uk).

ible with participant-centric design, but the rights and interests of participants are often subordinate or marginal, rather than incorporated as a central design principle. At the very least, the role of participants in many of these efforts is not being made visible.

The participant-centric model requires ongoing, meaningful efforts to engage the participants and potential participants; it is not reducible to one-off engagement events. Systematic efforts to gather input from participants can inform policy and practice, as well as guide specific projects. But they are not enough. Our research revealed three important design principles for ensuring participant-centricity in an MIC: (1) an MIC design should ideally involve opt-in

> In sum, participant-centricity has a robust ethical and pragmatic policy justification and is critical for commons sustainability. Yet, despite prevailing rhetoric about the importance of being participant-centric, a landscape analysis of data-sharing efforts revealed that there is considerable work to be done.

Some research study designs (for example, for most common disorders) would benefit from ready access to data about many people with few use restrictions, and simply cannot be organized around particular known diseases. Since all the uses cannot be foreseen in advance, the governance structure for such broadly used resources must rely on process and must credibly represent the rights and interests of the entire community. Attention also needs to be paid to privacy and security. There is some optimism that the technical solutions to allowing greater access to data while reducing privacy risks may emerge soon. Specifically, blockchain, homomorphic encryption and other modern multi-party encryption schemes may soon assist in enforcing process and policy; these ideas need to be pursued and, if workable, implemented.[21]

In sum, participant-centricity has a robust ethical and pragmatic policy justification and is critical for commons sustainability.[22] Yet, despite prevailing rhetoric about the importance of being participant-centric, a landscape analysis of data-sharing efforts revealed that there is considerable work to be done. The current system varies greatly in its involvement of participants and respect for their rights and interests. Many data structures have been built to address the needs of researchers, clinicians, private firms, or institutions, or were built in an era when individual control and reducing vulnerability to mass privacy breaches were not priorities; this is not necessarily incompat-

consent by individual participants, (2) meaningful, ongoing representation of participants in governance is essential, and (3) there needs to be dynamic interaction with participants over time.

The clear message from community members who participated in one of the three Community Advisory Panels (CAPs) and were asked to weigh in on policies for an MIC is that opt-in consent should be required for inclusion in an MIC.[23] This is consistent with other studies that have found that participants prefer opt-in consent, except in unusual cases of public health or other collective good, for which opt-out may be appropriate.[24] We need not revisit here the general arguments for and against carrying out some research in the absence of consent or with opt-out options. We acknowledge that weighing all relevant considerations, opt-out or no consent may be appropriate in contexts such as public health surveillance, use of de-identified information from a single source, like an electronic health record or newborn blood spot collection, or for research where an inclusive dataset is genuinely necessary in order to produce unbiased results, subject to careful oversight and accountability mechanisms.[25] We further acknowledge that the very flexibility we have highlighted as a desirable feature of a commons may lead to stakeholders (participants included) setting rules for particular data commons that do not involve opt-in consent. This may be especially likely in countries and health system contexts

where norms of solidarity and equality loom large and where there is less concern about potential insurance discrimination.[26] Nevertheless, during our research we found that MICs can trigger concerns that militate against opt-out or presumed consent.[27] Members of our advisory panels pointed to opt-out regimes in other contexts and described them as exploitative, noting they can erode trust and may come across as "sneaky." Many of them favored dynamic, granular consent as an alternative. However, some did express reservations about this approach, concerned about participation of the digitally unsophisticated, potential burdens on the research enterprise, and unwelcome intrusion from incessant consent requests. Thus, they seemed to settle on a more balanced approach that involves both opting-in to participation and maintaining some individual control over data uses, coupled with knowing

> **Perhaps one of the most important things we have learned from our work about participant-centricity is that if the rhetoric of reciprocal relationship (or participant empowerment, partnership, or engagement) is not matched by performance, distrust will result. One or a few instances of misuse of data or exploitation of participants have the potential to create skepticism or outright hostility toward the entire enterprise of building an MIC.**

how their data are being used and governance structures that ensure fair and meaningful representation.

There is also general support for the idea that participants should have a seat at the table when decisions are being made, and participant representatives should be selected through a process that is attuned to diversity.[28] When given a choice among several options, a clear majority of community members across sites favored participant representatives having voting rights on governing boards. In addition, community members emphasized the importance of incorporating diversity of participants in governance. The expert stakeholders we interviewed who are involved in various aspects of data-sharing initiatives from diverse employment sectors (i.e. laboratory, academia, non-government organization, government, technology, and healthcare company) also emphasized

the important role of participants in governance.[29] A few talked about this at some length, reporting dissatisfaction with the way that community representation on institutional review boards has worked out in practice and cautioning against tokenism. Some stressed the importance of giving participants a voice in choosing their representatives, for example via some kind of election process.[30] Expert stakeholders also understood and grappled with the importance of diversity. While a complete response to the diversity challenge was elusive, there was agreement about finding ways to involve a broad range of public representatives as well as advocacy communities and using a process for selection that results in geographical, socioeconomic, ethnic, health status, and other kinds of diversity. It was acknowledged that participants' rights and interests can only be fully represented if their diversity is recognized and represented.[31]

Finally, stakeholders from our research felt that it is important to interact with participants over time (but respect the wishes of those who would prefer not to be re-contacted). The individuals who participated on our community advisory panels were eager for a dynamic, reciprocal relationship, rather than a single transaction.[30] The professional stakeholders we interviewed — some of whom were themselves also participants in an MIC — mentioned that a few firms or platforms create an appealing participant experience. For example, professionals mentioned the Platform for Engaging Everyone Responsibly (PEER), which uses Private Access to allow individuals to tailor data sharing to their preferences and contribute data by completing self-paced survey instruments; the approach 23andMe has developed for reaching out to participants without being overly intrusive; and LunaDNA's proposed community-owned sharing model, which gives individuals who contribute their data for research a share in the profits generated from data access fees paid by researchers.[33] Community members and expert stakeholders also discussed the pros and cons of several strategies for giving information back to participants in an ongoing relationship with an MIC, including accessing raw data, returning individual and general results, showing community benefit, and offering opportunities to engage in citizen science. All are worth considering.

Perhaps one of the most important things we have learned from our work about participant-centricity is that if the rhetoric of reciprocal relationship (or par-

ticipant empowerment, partnership, or engagement) is not matched by performance, distrust will result. One or a few instances of misuse of data or exploitation of participants have the potential to create skepticism or outright hostility toward the entire enterprise of building an MIC. For an MIC to be successful, it must be trustworthy.

## How Can we Build Systems that are Trustworthy?

Several scholars have drawn attention to the importance of building trust among research participants for genetic studies.[34] Trust in the context of big data initiatives has been defined as "the willingness of a trustor to accept the potential risks involved in the sharing and further use of their personal data resulting from both optimism about the trustees' goodwill and interest in the public good."[35] Several of the expert stakeholders we interviewed and participants on our community advisory panels described how difficult it is to re-build trust once it has been eroded. Building and maintaining trust was viewed by many as one of the most difficult and most important non-technical challenges of building an MIC.

Taking seriously the commitment to make an MIC participant-centric can go a long way in building and maintaining trust. In particular, building trust was seen as an iterative process that goes well beyond informed consent. Engaging with participants and including them in every aspect of an MIC, including its governance, shows respect for the values and priorities of participants and emphasizes the good intentions of other stakeholders involved in building and maintaining an MIC. In addition, in order for participants to have warranted trust in an MIC, an MIC must prove to be trustworthy.[36] Both expert stakeholders and members of our community advisory panels described four critical attributes of a trustworthy MIC: transparency, access to data within the MIC, security, and accountability.

Transparent and truthful communication was viewed by members of our community advisory panels as an intrinsic moral obligation, something that is owed to participants in exchange for their agreeing to share information with an MIC.[37] Similarly, the GA4GH identified transparency, conceptualized as clear and accessible information on data-sharing practices, as one of the principal aspects of responsible data sharing.[38] Individuals we spoke with also recognized that the more transparent an MIC is, the more successful it will be in recruiting participants and obtaining access to their sensitive health and related information. Some felt that simply communicating truthful information to participants would be enough; others expressed the need for more engaged communication that builds relationships and forms collaborations. Participants were interested in transparent communication about how data are being used (or not) and why, how the commons is governed, what security mechanisms are in place, when data breaches occur, and how data breaches are being dealt with. Cultural change is needed in the research and clinical care communities so that transparent communication is woven into the fabric of all the data commons and other initiatives that make up an MIC. Technical solutions are also needed to facilitate more user-friendly and accessible modes of communication.

Giving participants access to their own data, and making sure that those data meet quality standards, are also considered essential elements of a trustworthy MIC. In the U.S., individuals have a right under the Health Insurance Portability and Accountability Act (HIPAA) to request a copy of their health data from covered entities. This right of access applies to all data about the individual held by a covered entity in one or more "designated record sets," and broadly includes medical records and "other records that are used, in whole or in part by or for the covered entity to make decisions about individuals."[39] Even when not legally compelled, however, giving participants access to their own data can empower them and engender trust.[40] Yet, several of the expert stakeholders we spoke with stated that people rarely request their data. If this is true, it is important to study why (e.g., do participants not know they can access their data and how to do it, or are they being obstructed in some way?).[41] Regardless of whether there is a legal right of access, MICs should assume an obligation to ensure that individuals have access to their individual-level data, if desired. If this principle is built into the design from the beginning, it will ensure the system retains this feature, which is essential to long-term trust. Trying to retrofit it into a system that has not accommodated individual access will be much more expensive and difficult.

Participants on our community advisory panels also emphasized the importance of having access to data that are accurate and reliable. Our expert stakeholders agreed, noting that much of the data currently in data commons are of such low quality that the data cannot be trusted.[42] Funders, data generators, and other stakeholders should thus ensure that quality standards are in place and must, at a minimum, be transparent about the quality of their data.[43] This requirement is important for research integrity and trustworthiness of the system; if participants can access their information in an MIC but that information is inaccurate or unreliable, then an MIC may be perceived as untrust-

worthy and efforts to build trust by providing access could be undermined.

Finally, in order for an MIC to be trustworthy it must have security measures in place to ensure that data are used only by approved users for authorized purposes. Data security is largely a technical challenge, one critical for building a trustworthy MIC. Recent privacy breaches, such as the ones at Equifax and Facebook, have increased concern about the security of personal, financial, and health information.[44] De-identification may be one way to protect privacy in the event of a data breach, although re-identification may be difficult to prevent. Some uses of data can be free and open, such as when "data altruists"[45] have deliberately made their data freely available, such as through Open Humans (https://www.openhumans.org), or when the private elements are not needed for analysis. However, most data in an MIC are much more valuable when they can be linked to individuals, and if there is sufficient data about an individual within an MIC, including genomic data, then there is a potential for re-identification.[46] For this reason, mechanisms beyond de-identification are needed to secure personal data, including having a robust informatics system that tracks approved uses and users of the information and prevents unauthorized access and misuse, including unauthorized downstream re-disclosure and re-use. In the event of a breach, participants should immediately be informed and told what will be done to protect them and to prevent a similar incident from happening in the future. The rules for governing data security as an important element of the commons are a work in progress.

A more robust system of accountability, with sanctions for misuse, is needed. Some participants from our community advisory panels worry about what legal recourse they have if their data are misused or stolen. The systems of accountability that are in place were uniformly thought to be inadequate because they are not comprehensive and there is insufficient enforcement and sanctions. For example, as mentioned above, although GINA protects against genetic discrimination in employment and health insurance in the U.S., it does not protect against discrimination in life, disability, and long-term care insurance.[43] Likewise, although there are sanctions imposed under HIPAA for the misuse of protected health information in the U.S., individuals have no real recourse when they become a victim of a privacy breach. There is no appeals process for IRB determinations, and there are limited enforcement mechanisms to ensure that institutions that share data do so according to the use limitations (the most severe sanctions typically consist of loss of data access privileges or funding, versus civil or criminal penalties). A more comprehensive system of accountability rules that has clear enforcement mechanisms in place, which could include audits, a private right of action for participants, or stronger sanctions for violations, would help build trust in the system.[48]

## Conclusion

This paper considers the potential for a sustainable MIC that facilitates open and responsible data sharing to emerge from the "ground up," through loose coordination and adherence to a set of high-level design principles rather than through imposition of a uniform structure. As Elinor Ostrom described, those high-level design principles must be agreed to by all stakeholders and should appeal to both ethics and pragmatism. Our research suggests that stakeholders broadly agree that an MIC should be both participant-centric and trustworthy. In addition, in order for the data resources that emerge to be sufficiently powerful to support the type of research needed to understand genotype-phenotype correlations, revolutionize clinical care, and reinvigorate public health, including addressing health disparities, custodians must be more concerned with how to share these resources openly yet responsibly than with maximizing profit or professional notoriety from the data. We also need to make sure that with so many different actors and rules in play, there are clear ways to create a networked environment that allows data to be easily aggregated and used. Otherwise, we risk the problem of an *anti*-commons: limited, cumbersome access and limited use.[49]

Despite these concerns, building an MIC is not a pipe dream. The gap between its achievement and its potential will depend critically on design decisions about participant-centricity and trustworthiness for both the constituent elements and for the system as a whole. Those design decisions lie in the hands of individual participants and their advocates, research institutions, governments, private firms and nonprofit organizations. We will all reap the benefits of the system we design.

## References

1. See R. Boyd, P.J. Richerson, R. Meizen-Dick, and T. De Moor et al., "Tragedy Revisited," *Science* 362, no. 6420 (2018): 1236-1241.
2. National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (Washington, D.C.: The National Academies Press, 2011).
3. R. Cook-Deegan, M.A. Majumder, A.L. McGuire, "Introduction: Sharing Data in a Medical Information Commons," *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 7-11; P.A. Deverka, M.A. Majumder, A.G. Villanueva and M. Anderson et al., "Creating a Data Resource: What Will It Take to Build a Medical Information Commons?" *Genome Medicine* 9, no. 84 (2017): 1-5, *available at* <https://genomemedicine.biomed-central.com/articles/10.1186/s13073-017-0476-3> (last visited January 4, 2019).
4. See C. Hess and E. Ostrom, eds., *Understanding Knowledge as a Commons: From Theory to Practice* (Cambridge; London: MIT Press, 2011); E. Ostrom, *Governing the Commons* (Cambridge: Cambridge University Press, 1990); and E. Ostrom, *Understanding Institutional Diversity* (Princeton: Princeton University Press, 2005).
5. M.A. Majumder, J.M. Bollinger, A. G. Villanueva, P.A. Deverka, and B.A. Koenig, "The Role of Participants in a Medical Information Commons," *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 51-61.
6. P.A. Deverka, D. Gilmore, J. Richmond, and Z. Smith et al., "Hopeful and Concerned: Public Input on Building a Trustworthy Medical Information Commons," *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 70-87.
7. N. Anderson, C. Bragg, A. Hartzler, and K. Edwards, "Participant-Centric Initiatives: Tools to Facilitate Engagement in Research," *Applied & Translational Genomics* 1 (2012): at 25; J. Kayne, L. Curren, N. Anderson, and K. Edwards et al., "From Patients to Partners: Participant-Centric Initiatives in Biomedical Research," *Nature Reviews Genetics* 13 (2012): 371-376.
8. M.P. Ball, J.R. Bobe, M.F. Chou, and T. Clegg et al., "Harvard Personal Genome Project: Lessons from Participatory Public Research," *Genome Medicine* 6, no. 10 (2014).
9. Ostrom's case studies of successful commons typically involved activity on a relatively small scale with participants who shared values and goals and had ongoing relationships with one another. At the same time, Ostrom and colleagues articulated a principle of "nesting" that allows for governance at multiple levels. T. Dietz, E. Ostrom, and P.C. Stern, "The Struggle to Govern the Commons," *Science* 3012 (2003): 1907-1912. In this respect, the vision we articulate of an MIC as a collection of linked data commons that vary along a number of dimensions but all adhere to a set of high-level design principles (based on multi-stakeholder input) in arriving at their particular rules aligns with Ostrom's work. See M.A. Majumder, P.D. Zuk, A.L. McGuire, "Medical Information Commons," in B. Hudson, J. Rosenbloom, and D. Cole, eds., *Routledge Handbook of the Study of the Commons* (Routledge, Forthcoming 2019).
10. For example, the concept of "inferential disclosure," which refers to the potential use of available data to determine the value of some characteristic of an individual more accurately than would have been possible absent that data, captures the risk of harm that exists even absent definitive re-identification of an individual's information within a dataset. See, e.g., "Report of the Committee on National Statistics' Panel on Confidentiality and Data Access," G.T. Duncan, T.B. Jabine, and V.A. de Wolf, eds., *Private Lives and Public Policies* (Washington, DC: National Academy Press, 1993): at 24.
11. See S.U. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018); C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown Publishing Group, 2016); V. Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York: St. Martin's Press, 2018).
12. See National Institutes of Health, *Request for Comments: Proposal to Update Data Management of Genomic Summary Results Under the NIH Genomic Data Sharing Policy*, National Institutes of Health Website, *available at* <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-17-110.html> (last visited January 4, 2019); S. O. Dyke, E.S. Dove, and B.M. Knoppers, "Sharing Health-Related Data: A Privacy Test?" *Genomic Medicine* 1, Article no. 16024 (2016).
13. R.C. Green, D. Lautenbach, and A.L. McGuire, "Gina, Genetic Discrimination, and Genomic Medicine," *New England Journal of Medicine* 372, no. 5 (2015): 397-399; M.A. Rothstein, "GINA at Ten and the Future of Genetic Nondiscrimination Law," *Hasting Center Report* 48, no. 3 (2018): 5-7.
14. Y. Erlich, J.B. Williams, D. Glazer, and K Yocum et al., "Redefining Genomic Privacy: Trust and Empowerment," *PLOS Biology* 12, no. 11 (2014): e1001983, *available at* <https://doi.org/10.1371/journal.pbio.1001983> (last visited January 4, 2019).
15. S. Scheridan, S. Schrandt, and L. Forsythe, Advisory Panel on Patient Engagement, T. Hilliard, K. and A. Paez, "The PCORI Engagement Rubric: Promising Practices for Partnering in Research," *Annals of Family Medicine* 15, no. 2 (2017): 165-170; L.E. Ellis and N.E. Kass, "How Are PCORI-funded Researchers Engaging Patients in Research and What Are the Ethical Implications?" *AJOB Empirical Bioethics* 8, no. 1 (2017): 1-10; E. Nelson, M. Dixon-Woods, P.B. Batalden, K. Homa, et al., "Patient-Focused Registries Can Improve Health, Care, and Science," *BMJ* 354 (2016).
16. PE International, PXE International Facebook group webpage, *available at* <https://www.facebook.com/groups/PXEers/> (last visited January 4, 2019); S. Lozinsky, Chordoma Connection is here! (January 23, 2018), Chordoma Foundation Website, *available at* <https://www.chordoma-foundation.org/latest-updates/chordoma-connections-is-here/> (last visited January 4, 2019).
17. The Life Raft Group, "The LRG Mission & Vision," The LRG Website, *available at* <https://liferaftgroup.org/the-lrg-mission-vision/> (last visited January 4, 2019).
18. See the Home and Research Results pages on the Multiple Myeloma Research Foundation Website, *available at* <https://themmrf.org/> (last visited January 4, 2019).
19. M.P. Feldman and A. Graddy-Reed, "Accelerating Commercialization: A New Model of Strategic Foundation Funding," *Journal of Technology Transfer* 39, no. 4 (2014): 503–523, *available at* <https://doi.org/10.1007/s10961-013-9311-1> (last visited January 4, 2019).
20. Another example is the David Fajgenbaum and Castleman Disease Collaborative Network. See D.C. Fajgenbaum, J.R. Ruth, D. Kelleher, and A.H. Rubenstein, "The Collaborative Network Approach: A New Framework to Accelerate Castleman's Disease and Other Rare Disease Research," *The Lancet Haematology* 3, no. 4 (2016): PE150-E152.
21. Blockchain is described by Ozercan et al. (2018) as a distributed database technology. In the context of an MIC, blockchain would allow for decentralized MIC comprised

of multiple databases. For a discussion of blockchain and genomic data sharing, including examples of emerging data-sharing projects, see H.I. Ozercan, A.M. Ileri, E. Ayday, and C. Alkan, "Realizing the Potential of Blockchain Technologies in Genomics," *Genome Research* 28, no. 9 (2018): 1255–1263, *available at* <https://doi.org/10.1101/gr.207464.116> (last visited January 4, 2019).

22. See Deverka, *supra* note 6.
23. See *id*.
24. D. Goodman, D. Bowen, P. Tehrani, and F. Fernando et al., "The Research Participant Perspective Related to the Conduct of Genomic Cohort Studies: A Systematic Review of the Quantitative Literature," *Translational Behavioral Medicine* 8, no. 1 (2018): 119-129 (focus on large genomic cohort studies); K.K. Kim and L. Ohno-Machado, "Comparison of Consumers' Views on Electronic Data Sharing for Healthcare and Research," *Journal of the American Medical Informatics Association* 22, no. 4 (2015): 821-830 (focus on electronic data use during an emergency). See also D. Goodman, C.O. Johnson, L. Wenzel, and D. Bowen, "Consent Issues in Genetic Research: Views of Research Participants," *Public Health Genomics* 19, no. 4 (2016): 220-228; S.C. Hull et al., "Patients' Views on Identifiability of Samples and Informed Consent for Genetic Research," *American Journal of Bioethics* 8, no. 10 (2008): 62-70. But see N.A. Garrison et al., "A Systematic Literature Review of Individuals' Perspectives on Broad Consent and Data Sharing in the United States," *Genetics in Medicine* 18, no. 7 (2016): 663-671.
25. See National Academies of Sciences, Engineering, "Legal and Computer Science Approaches to Privacy," in R.W. Groves and B.A. Harris-Kojetin, eds., in *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps* (Washington, D.C., The National Academies Press, 2017): 61-78, *available at* <https://doi.org/10.17226/24893> (last visited January 4, 2019; J.R. Botkin, et al., "Retention and Research Use of Residual Newborn Screening Bloodspots," *Pediatrics* 131, no. 1 (2013): 120–27; J. Illman, "Cancer Registries: Should Informed Consent Be Required?" *JNCI: Journal of the National Cancer Institute* 94, no. 17 (2002): 1269–1270.
26. For example, in the UK, discussion of sharing clinical (genomic) data has been framed in terms of a social contract. See annual report of the Chief Medical Officer 2016: Generation Genome, *available at* <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/631043/CMO_annual_report_generation_genome.pdf> (last visited January 4, 2019).
27. See Deverka, *supra* note 6.
28. Several seats refers to representation at multiple levels of governance (e.g. steering committee, data access committee, etc.) and having multiple representatives on the same level (i.e. more than one on a data access committee to facilitate diversity).
29. J.M. Bollinger et al., "What Is a Medical Information Commons?" *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 41-50.
30. See Majumder, *supra* note 5.
31. Concerning the challenges of ensuring that patient representatives are truly "representative," see B. von Tigerstrom, "The Patient's Voice: Patient Involvement in Medical Product Regulation," *Medical Law International* 16, no. 1-2 (2016): 27-57; See Majumder, *supra* note 5.
32. See Deverka, *supra* note 6.
33. A.M. Li and S.F. Terry, "Linking Personal Health Data to Genomic Research," *Genetic Testing and Molecular Bio-*

*markers* 19, no. 1 (2014): 1–2; "LunaDNA - A Community Owned Genomic and Medical Research Platform," *available at* <https://www.LunaDNA.com/> (last visited January 4, 2019).
34. See J.J. Arias et al., "Trust, Vulnerable Populations, and Genetic Data Sharing," *Journal of Law and the Biosciences* 2, no. 3 (2015): 747-753.
35. A. Adjekum, M. Ienca, and E. Vayena, "What Is Trust? Ethics and Risk Governance in Precision Medicine and Predictive Analytics," *Omics: A Journal of Integrative Biology* 21, no. 12 (2017): at 706, *available at* <https://www.liebertpub.com/doi/pdf/10.1089/omi.2017.0156> (last visited January 4, 2019).
36. See *Id.*, at 705.
37. See Deverka, *supra* note 6.
38. B.M. Knoppers, "Framework for Responsible Sharing of Genomic and Health-Related Data," *Hugo Journal* 8, no. 1 (2014): 3, *available at* <https://thehugojournal.springeropen.com/articles/10.1186/s11568-014-0003-1> (last visited January 4, 2019).
39. Code of Federal Regulations, Title 45, Parts 160, 164. 38. 40. A.
40. A. Thorogood, J. Bobe, B. Prainsack, and A. Middleton et al., "APPLaUD: Access for Patients and Participants to Individual Level Uninterpreted Genomic Data," *Human Genomics* 12, no. 7 (2018), *available at* <https:// humgenomics.biomedcentral.com/articles/10.1186/ s40246-018-0139-5> (last visited February 21, 2019).
41. See GetMyHealthData.org for stories about how patients have faced challenges in requesting access to their data.
42. See Bollinger, *supra* note 29.
43. J. R. Botkin, M. Mancher, E.R. Busta, and A.S. Downey, eds, *Returning Individual Research Results to Participants: Guidance for a New Research Paradigm* (July 2018), *available at* <http://www.nationalacademies.org/hmd/Reports/2018/returning-individual-research-results-to-participants.aspx> (last visited January 4, 2019).
44. S. Gressin, *The Equifax Data Breach: What to Do* (September 2017), Federal Trade Commission Website, *available at* <https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-doK> (last visited January 4, 2019); K. Fazzini and C. Farr, *Facebook "Closed" Groups Weren't As Confidential as Some Thought* (August 2018), CNBC Website, *available at* <https://www.cnbc.com/2018/07/11/facebook-private-groups-breast-cancer-privacy-loophole.html> (last visited January 4, 2019).
45. I.S. Kohane and R.B. Altman, "Health Information Altruists – A Potentially Critical Resource," *New England Journal of Medicine* 353, no. 19 (2005): 2074-2077.
46. M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, and Y. Erlich, "Identifying Personal Genomes by Surname Inference," *Science* 339 (2013): 321-324.
47. The Genetic Information Nondiscrimination Act of 2008, Pub. L. 110-233, 122 Stat. 881 (2008), *available at* <https://www.gpo.gov/fdsys/pkg/PLAW-110publ233/html/PLAW-110publ233.htm> (last visited January 4, 2019).
48. M. Phillips, E.S. Dove, and B.M. Knoppers, "Criminal Prohibition of Wrongful Re-identification: Legal Solution or Minefield for Big Data?" *Bioethical Inquiry* 14, no. 4 (2017): 527–539.
49. See discussion of a possible anti-commons in the context of biomedical research in M.A. Heller and R.S. Eisenberg, "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *Science* 280, no. 5364 (1998): 698–701.