Genomic Data-Sharing Practices

Angela G. Villanueva, Robert Cook-Deegan, Jill O. Robinson, Amy L. McGuire, and Mary A. Majumder

Additional materials related to this article are available online at journals.sagepub.com/home/lme

Making human data in research and clinical datasets broadly accessible is essential to advancing biomedical research, precision medicine, and public health. Research data are often shared in compliance with policies, such as those set forth by funding agencies (e.g., the National Institutes of Health (NIH) Genomic Data Sharing Policy) and by academic journals (e.g., in compliance with the International Committee of Medical Journals Editors).[1] Advocates and supporters of open science are also encouraging sharing of datasets and other valuable resources that enable data integration and analysis.[2] Not only are data being shared, but institutional arrangements are being developed to support the distribution of data, creating a medical information commons (MIC).[3] An MIC is a networked environment in which diverse health, medical, and genomic data on large populations become widely shared resources.[4] While this paper uses the term "MIC" to refer to the data-sharing ecosystem, certain data-sharing efforts within the ecosystem may also be described as MICs. Such an understanding of an MIC follows Elinor Ostrom and colleagues' conceptualization of commons for managing larger common-pool resources as complex and having multiple layers.[5]

Policies and guidelines endorsed by organizations such as the Global Alliance for Genomics and Health (GA4GH) and the Organisation for Economic Cooperation and Development (OECD) promote the development of an MIC and recommend that data-sharing initiatives communicate information on key features such as governance, privacy and security protections, and data access rules in a transparent manner and through accessible formats, including digital platforms.[6] Transparency about data-sharing practices can cultivate trust among individuals con-

Angela G. Villanueva, M.P.H., *is a Research Associate at the Center for Medical Ethics and Health Policy at Baylor College of Medicine.* **Robert Cook-Deegan, M.D.**, *is a Professor in the School for the Future of Innovation in Society at Arizona State University. He is a physician and molecular biologist who turned to policy and then entered academe through Georgetown, Stanford, and Duke Universities before joining ASU.* **Jill O. Robinson, M.A.**, *is the Research Manager at the Center for Medical Ethics and Health Policy, Baylor College of Medicine. She received her B.A. in sociology and political science and her M.A. in sociology from the University of Houston.* **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.*

The Journal of Law, Medicine & Ethics, 47 (2019): 31-40. © 2019 The Author(s)
DOI: 10.1177/1073110519840482
https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

templating participation in or considering whether to continue their participation in a data-sharing initiative.[7] Achieving a trusted and trustworthy MIC is imperative if the benefits anticipated from large-scale data sharing are to be realized.[8] But how well are data-sharing initiatives living up to this norm of transparency about their practices? And what can be learned from comparing practices across data-sharing initiatives, insofar as those practices can be identified drawing on available information?

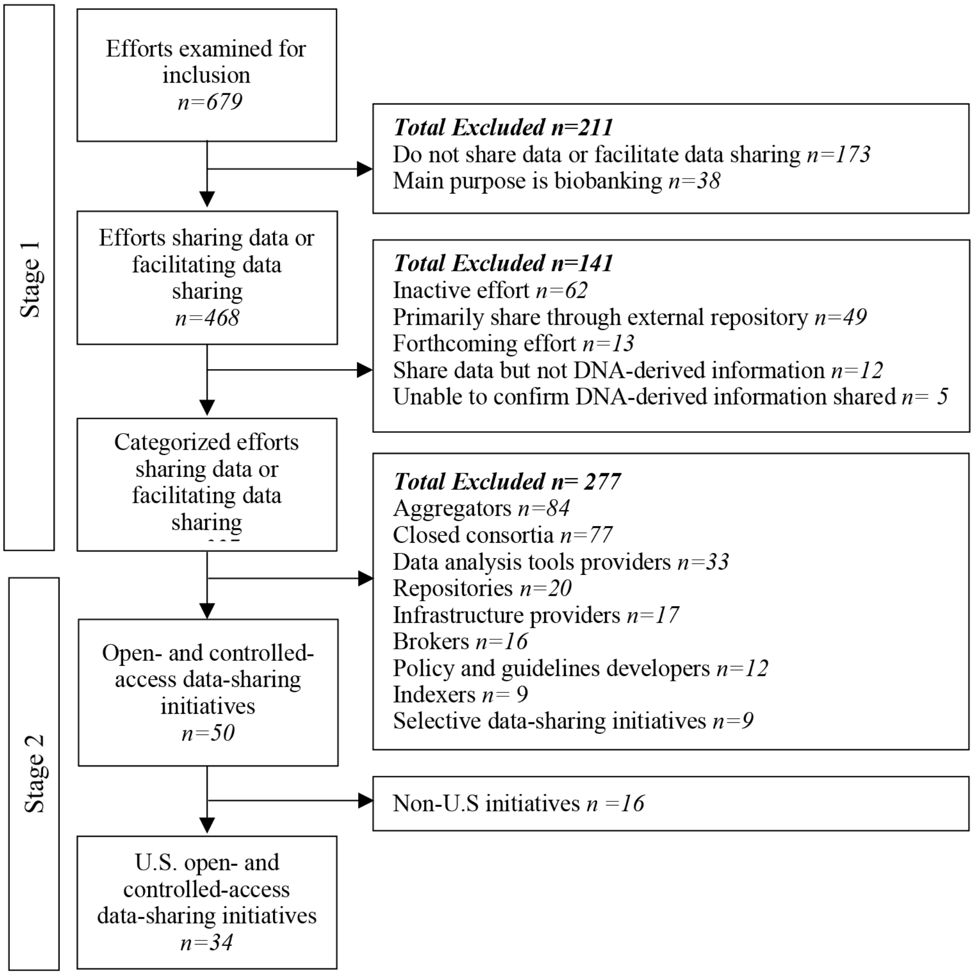
We reviewed publicly-available information about existing data-sharing initiatives in order to comment on the implementation of strategies recommended by GA4GH and other organizations to promote transparency about data-sharing practices. Our focus was on U.S. efforts collecting and sharing primary DNA-derived data from individuals, and we report on characteristics such as funding source and type(s) of data shared and practices in the domains of consent, privacy and security, data access, oversight, and participant engagement.

Methods

A two-stage analysis was carried out to describe the biomedical data-sharing landscape. To inform current research efforts, including precision medicine projects, the landscape analysis focused on efforts sharing or facilitating the distribution of genomic

Figure 1

Landscape Analysis Consort Diagram


Stage 1
Stage 2
Efforts examined for inclusion *n=679*
Efforts sharing data or facilitating data sharing *n=468*
Categorized efforts sharing data or facilitating data sharing
Open- and controlled-access data-sharing initiatives *n=50*
U.S. open- and controlled-access data-sharing initiatives *n=34*
Total Excluded n=211
Do not share data or facilitate data sharing *n=173*
Main purpose is biobanking *n=38*
Total Excluded n=141
Inactive effort *n=62*
Primarily share through external repository *n=49*
Forthcoming effort *n=13*
Share data but not DNA-derived information *n=12*
Unable to confirm DNA-derived information shared *n= 5*
Total Excluded n= 277
Aggregators *n=84*
Closed consortia *n=77*
Data analysis tools providers *n=33*
Repositories *n=20*
Infrastructure providers *n=17*
Brokers *n=16*
Policy and guidelines developers *n=12*
Indexers *n=* 9
Selective data-sharing initiatives *n=9*
Non-U.S initiatives *n =16*

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

data that share or do not share other health-related information. The first stage of the analysis examined publicly-available online information and led to the development of a typology characterizing the data-sharing landscape as a whole. Details regarding the first stage of our landscape analysis are reported elsewhere in this issue.[9] In brief, we reviewed a total of 679 efforts. After application of exclusion criteria as shown in Figure 1, the remaining 327 were used to generate a typology of efforts sharing or facilitating the distribution of data derived from human DNA.[10] Note that at this stage we excluded efforts that share data only through an external repository. While efforts contributing data to repositories (such as the database of Genotypes and Phenotypes (dbGaP)) are important to the development and sustainability of an MIC, data shared through repositories are governed by the policies established by the repository. Biobanks only sharing biosamples (i.e., not sharing data) and forthcoming initiatives (i.e., initiatives in the planning stages) were also excluded, along with efforts no longer actively sharing data.

Stage Two of Landscape Analysis

The second stage of the landscape analysis reports on the data-sharing practices of the subset of efforts that were categorized in the first-stage analysis as open- and controlled-access data-sharing initiatives.[11] These initiatives were chosen for further review because they directly collect from individuals biosamples used to derive genetic data or collect reports containing genetic data such as the reports furnished by direct-to-consumer DNA testing companies, allowing for an examination of practices from the point of consent to the distribution of the data. The open- and controlled-access data-sharing initiatives included for review satisfied the following inclusion criteria: 1) share DNA-derived data generated from biosamples or data reports collected from individuals; 2) are of U.S. origin or include data collected in the U.S.; and 3) share the data through distinct mechanisms, meaning according to practices established by the initiative itself. Finally, initiatives that had insufficient publicly-available information about data-sharing practices to enable characterization of those practices in any of our major domains (Table 1) were excluded, yielding a total of 34 data-sharing initiatives reviewed for purposes of this stage of our analysis.

Data Collection and Analysis

To comment on transparency, our review exclusively relied on publicly-available online information, including supporting documents (e.g., consent forms, data-sharing agreements, data dictionaries) if available. Information and documents available only upon request or following completion of a registration process were not included. Publicly-available, Internet-accessible academic and popular press articles were reviewed to better understand the operations of a given initiative. Details on data-sharing practices were captured on a data collection instrument developed by the research team and programmed into the Research and Electronic Data Capture (REDCap) platform.[12] The process of developing the data collection instrument was iterative and initially guided by a preliminary review of websites and project team input. The final instrument included questions of interest and response categories informed by our review of initiatives and relevant guidelines on data sharing and transparency. Initiatives included in the final set for analysis did not necessarily have details related to all the Table 1 domains available on their website; we therefore used a "Not Mentioned Online" response category to capture these occurrences. The REDCap data collection instrument is available as part of the online Supplemental Information for this paper.

Each website and related documents were examined for various characteristics (Table 1) including funding source, type of data shared, and promoted use of the data (e.g., disease-specific vs. broad data use). We categorized funding from academic cen-

Table 1
Reported Information on Characteristics and Data-Sharing Practices

Characteristics	Funding source Type(s) of data shared Adult vs. pediatric data shared Disease-specific or broad focus
Data-Sharing Practices:	
Consent	Type of consent
Privacy and Security	Discussion of applicable privacy laws Breach response plan
Data Access	Data access levels Requirements to access data Invited data users International data sharing
Oversight	Data access oversight body Constitution of data access oversight body
Participant Engagement	Mode(s) of communication with participants Engagement of participants in decision making

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

ters as non-profit funding. Although we focused on data-sharing initiatives sharing DNA-derived data, we recorded whether other types of data that are important factors in health, such as the built environment and lifestyle habits, were also included in shared datasets. We also explored practices related to consent, privacy and security, data access, oversight, and participant engagement. For privacy, we examined whether initiatives had standalone webpages or downloadable documents and sections on a webpage dedicated to sharing information about data privacy protections. We also took note of any information related to privacy posted elsewhere, such as on the consent form or the Terms of Use (TOU) agreement. TOU agreements typically serve to inform the user of the utilization of HTTP cookies to collect browsing activity data, in addition to other information governing the use of a website. We only included in our review TOU agreements that described policies and practices related to the sharing of health-related data. In addition, we examined the accessibility of data and used the GA4GH definition of open access — "[m]aking data available without restriction" — to classify an initiative as open- or controlled-access.[13] We considered any gatekeeping action required to access the data– for example, submitting an application or creating a user account– a restriction, and such initiatives were classified as controlled-access.

Also, engagement of individuals whom the data describe ("participants") was examined in two ways: communication with participants through the initiative's website and the involvement of participants in making decisions concerning data uses. All the initiatives reviewed had a public interface. However, we distinguished between websites intended for researchers and websites intended for participants and did not consider posting content intended for researchers, such as study protocols, a form of participant engagement.

Two members of the research team archived copies of digital content, entered data, and reviewed each other's entries. We started the landscape analysis in 2015 and continuously updated the information through August 2018. The use of REDCap allows for an automatic assignment of numerical codes for the response options. The reported relative frequencies were derived from REDCap using the Reports function.

Results

A total of 34 initiatives that met inclusion criteria of sharing DNA-derived data of U.S. origin or with significant U.S. presence and publicly available data-sharing details are included in this analysis.[14]

Characteristics

More than half were funded by non-profit, non-governmental organizations (n=19, 55.9%), such as academic centers, advocacy organizations, and private foundations. Other funding sources included governmental organizations (n=6, 17.6%) and for-profit companies (n=2, 5.9%), in addition to partnerships between governmental, non-profit, and for-profit entities (n=3, 8.8%), government and non-profit organizations (n=2, 5.9%), and non-profit organizations and for-profit companies (n=2, 5.9%). Most initiatives were established by a U.S.-based organization (n=31, 91.2%) and promoted disease-specific research (n=25, 73.5%).

The vast majority of initiatives (n=33, 97.1%) shared some phenotypic information (e.g., disease diagnosis, weight, height, handedness), and several also shared other health-related information, such as diet (n=11, 32.4%), physical activity (n=12, 35.3%), and drug (including cigarettes) and alcohol use (n=15, 44.1%). Other data shared included information on education (n=14, 41.2%), employment status and income (n=12, 35.3%), the built environment, including zip code (n=10, 29.4%), health insurance status (n=5, 14.7%), and social connections (n=4, 11.8%).[15] Nearly half shared data from both adults and minors (n=15, 44.1%), and a third shared data from adults only (n=13, 38.2%), with the remainder not specifying online the age of participants represented in the data shared (n=6, 17.6%).[16]

Consent

We found information about consenting processes, consent forms, or TOU Agreements containing language typically seen in consent forms for 18 initiatives. One TOU Agreement, that of the Altruist Database, advised participants that it is their responsibility to review the TOU Agreement periodically to monitor for changes that may affect their willingness to participate. Most of the initiatives used a one-time agreement or broad consent (n=13, 72.2%). Only one initiative, the HEROIC Registry website, clearly described the use of a dynamic, granular consent process that that allows participants to specify the acceptable uses of the data they contribute and change these parameters over time. Specifically, the Registry linked to a Terms of Use pop-up window that disclosed their implementation of the Platform for Engaging Everyone Responsibly (PEER).[17]

Privacy and Security

Initiatives communicated information about data privacy via a standalone webpage or within a section of a webpage such as a Frequently Asked Questions

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

page (n=6, 17.6%), via a supporting document such as a consent form inclusive of a TOU agreement operating as a consent (n= 6, 17.6%), or via both webpage content and a supporting document (n=12, 35.3%). For 10 (29.4%) initiatives, we were unable to locate any information about data privacy and security. An example of an initiative communicating risks and privacy information via a TOU agreement was openSNP, which posted a disclaimer on the lack of data privacy.[18] Six initiatives (17.6%) noted the possibility of a data breach, while one (2.9%) indicated that they will notify participants of a data breach. Also, ten (29.4%) mentioned the Health Insurance Portability and Accountability Act (HIPAA) and eight (23.5%) mentioned the Genetic Information Nondiscrimination Act. Additional information reported via websites include dreceiving IRB approval (n=11, 32.4%) and receiving a Certificate of Confidentiality (n=8, 23.5%). A Certificate of Confidentiality would prohibit those initiatives from disclosing sensitive, identifiable information with only a few exceptions (including disclosure with individual consent and for research purposes under certain conditions).[19] An interesting data privacy and security strategy we recorded has been implemented by the Personalized Medicine Research Project at the Marshfield Clinic. The Clinic reported storing research data in a database that is not connected from "other Clinic information systems or to any external network, such as the Internet."[20]

Data Access

The majority of initiatives (n=28, 82.4%) shared data only through a controlled-access portal that requiresdsome action to be taken before data could be accessed (see Table 2). Three (8.8%) initiatives had no restrictions on data access and thus met the GA4GH definition of "open access," and three (8.8%) comprised both open- and controlled-access data. Some initiatives (n=4, 11.8%) made data available through an external repository and through distinct, initiative-level mechanisms. For example, the Framingham Heart Study (FHS) reported sharing genetic data through dbGaP; data not available through dbGaP were directly shared with researchers approved by the FHS DNA Committee.[21] We observed that 17 (50%) initiatives used the language of "open science" to describe their data-sharing policy.

The majority of initiatives (n=31, 91.2%) invited researchers to access data without specifying whether the researcher must be affiliated with an academic

Table 2
Requirements and Users by Type of Access

	Initiative Type			
	All Initiatives n= 34 (100%)	**Controlled-Access n= 28 (82.4%)**	**Controlled- & Open-Access♦ n= 3 (8.8%)**	**Open-Access n= 3 (8.8%)**
Requirements to Access Data ▲				
Agree to Conditions of Data Use	22 (64.7 %)	19 (67.9%)	3 (100%)	--
Create User Account	19 (55.9%)	16 (57.1%)	3 (100%)	--
Submit Application	16 (47.1%)	15 (53.6%)	1 (33.3%)	--
Obtain IRB Approval	14 (41.2%)	13 (46.4%)	1 (33.3%)	--
Send Email Request	8 (23.5%)	8 (28.6%)	0 (0%)	--
Access External Repository	4 (11.8%)	3 (10.7%)	1 (33.3%)	--
Invited Data Users ▲				
Unspecified Researchers	31 (91.2%)	26 (92.9%)	2 (66.7%)	3 (100%)
Academic Researchers	2 (5.9%)	1 (3.6%)	1 (33.3%)	0 (0%)
Private Industry Researchers	3 (8.8%)	2 (7.1%)	1 (33.3%)	0 (0%)
Clinicians/Healthcare Providers	5 (14.7%)	4 (14.3%)	1 (33.3%)	0 (0%)
Other ■	6 (17.6%)	3 (10.7%)	2 (66.7%)	1 (33.3%)
Information on who has accessed data is available online.	10 (29.4%)	9 (32.1%)	1 (33.3%)	--

▲ The column percent totals may sum to more than one hundred as more than one option may have been selected for each initiative.
♦ Information on requirements to access data pertains to data available through controlled access only.
■ "Other" category includes DTC genetic testing company customers and the public.

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

institution or other sectors, such as private industry. Healthcare providers and clinicians were identified as data users by six (17.6%) initiatives. Lastly, five (14.7%) initiatives explicitly invited other groups, for example, customers of direct-to-consumer (DTC) genetic testing companies or the general public, to access data. Categories of invited data users were not mutually exclusive.

In order to access data, the majority of initiatives reviewed (n=19, 55.9%) required data seekers to create a user account. Other common requirements include submitting an application (n=16, 47.1%) and/or receiving IRB approval (n=14, 41.2%). Furthermore, the majority of initiatives (n=22, 64.7%) required agreement to abide by stipulations governing data access, such as permissible data uses and acknowledgement of the data source in publications resulting from data accessed. Many of these initiatives required a signed agreement; a few used clickwrap agreements. An initiative granting data access through a clickwrap agreement was the American Association for Cancer Research Project Genie, which required data seekers to click on the "Agree" box following individual statements specifying the terms of data use and then click a 'SUBMIT" button. Thereafter a Google email address was used to authenticate the user before access was granted.[22] In other cases, applications for access underwent a review process to determine eligibility based on institutional affiliation and purpose of the request. For example, the Alzheimer's Disease Neuroimaging Initiative's Data Sharing and Publication Policy stated: "The DPC [Data and Publications Committee] does not believe it is feasible to adjudicate who is 'qualified' and who is not, however to maintain compliance with language in the informed consent documents indicating that data will be shared with members of the 'scientific community,' the DPC will review the applications of each investigator requesting data and make a judgment based on their affiliation with a scientific or educational institution, and on the basis of the reason for the request."[23]

Of the initiatives we reviewed, ten (29.4%) provided information on who has accessed data either by listing the publications resulting from data access or providing statistics on the sectors that have used the data. Furthermore, for most of the initiatives reviewed (n=25, 73.5%), no information on data release embargo policies was found; a few mentioned data release policies (n=9, 29%) that either state data are readily available (n= 7, 77.8%) or described the conditions of a data release embargo (n=2, 22.2%). For example, the Simons Foundation Autism Research Initiative embargo policy stated: "In order to facilitate other research projects…Simons Foundation personnel may, at their discretion, release these data to other qualified investigators with the understanding that these other investigators will have agreed not to publish on these data until after an embargo period expires."[24]

Information on fees to access data was also limited, as half of the initiatives did not post such information. Among the initiatives that disclosed information on fees (n=17, 50%), most (n=11, 64.7%) made data freely available. Additional sharing practices that were publicly available for a subset of the initiatives included also sharing data through external data resources, such as dbGaP (n=17, 50%) and sharing data with researchers outside the U.S. (n=22, 64.7%).

Oversight

Of the 31 initiatives that shared data through a controlled-access mechanism, 23 (74.2%) explicitly mentioned an oversight body, such as a data access committee, that reviews and approves data requests. Ten of those (n=10, 43.5%) publicized the roster of names and/or roles of the individuals involved with data-sharing decisions. Representatives with an academic affiliation were the most commonly mentioned members of data-sharing oversight bodies (n=9, 90%); however, some did mention including industry representatives (n=2, 20%) and participant representatives (n=4, 40%).[25]

Participant Engagement

The majority of the initiatives we reviewed (n=23, 67.6%) had websites containing content intended for participants that was communicated through electronic newsletters (n=13, 56.5%) or social media such as a Facebook page (n=5, 21.7%). In addition, some of the initiatives engaged participants in decision-making by allowing them to exercise some control over how their individual data are used (n= 3, 8.8%) and by whom (n=4, 11.8%), giving them the ability to edit or update their data (n=7, 20.6%), and/or involving them on an oversight or other governance body, such as a data access committee. For example, the Jackson Heart Study (JHS) was notable for its work in engaging participants. Through a Community Outreach Center (CORC), JHS engaged with participants and community members — most of whom are African American — to explain data uses, to promote study participation and retention, and to train community health advisors who in turn educate those living in surrounding communities about cardiovascular disease prevention.[26]

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

Discussion

As data-sharing initiatives continue to emerge and contribute to the evolving MIC, the time is opportune for a discussion of data-sharing practices and the transparency of these practices. Our analysis offers a snapshot of how existing U.S.-based initiatives that collect and share genomic and other health-related data present their data-sharing practices in several important domains. Here we highlight several areas in which practices appear to vary or merit further scrutiny. In particular, we focus on collection and sharing of environmental exposure (exposome) data, responses to growing concerns about privacy and security, data access requirements and citizen science, and the relationship between consent, oversight, and participant engagement.

As data-sharing initiatives continue to emerge and contribute to the evolving MIC, the time is opportune for a discussion of data-sharing practices and the transparency of these practices. Our analysis offers a snapshot of how existing U.S.-based initiatives that collect and share genomic and other health-related data present their data-sharing practices in several important domains. Here we highlight several areas in which practices appear to vary or merit further scrutiny. In particular, we focus on collection and sharing of environmental exposure (exposome) data, responses to growing concerns about privacy and security, data access requirements and citizen science, and the relationship between consent, oversight, and participant engagement.

Although we limited our review to initiatives that include primary DNA-derived data in the datasets they created and shared, we recognize the importance of populating an MIC with diverse types of health-related data to better understand common and complex disease and health outcomes.[27] Indeed, a 2011 U.S. National Academies report emphasized the exposome, stating that with such data as part of its foundation "a Knowledge Network of Disease could lead to better understanding of the variables and mechanisms underlying disease and health disparities, thereby helping to reveal a truer picture of the ecology of human health and facilitating a more holistic approach to health promotion and disease prevention."[28] Accordingly, efforts to share DNA-derived data should endeavor to incorporate and link more information related to physical and social environmental exposures in order to achieve the National Academies' vision and precision medicine and public health goals.

Amassing a variety of data about individuals, particularly genetic or sensitive behavioral primary data, raises privacy concerns even when the data are de-identified in line with current standards (e.g., as established in connection with HIPAA). Many of the initiatives we reviewed did not describe security measures or a breach response plan on their websites, although these initiatives may disclose that information through other media. We suggest that data-sharing initiatives adopt a practice of publicizing measures they are taking to prevent breaches and sharing information about their plan for communication with participants should a breach occur, especially given public attention to cyberattacks on services such as Equifax and Healthcare.gov.[29] Genomic data are not immune to hacking. In fact, the United Kingdom's National Health Service stores DNA data acquired for the Genomics England project on military servers due to data security concerns.[30] Similarly, initiatives that do not connect a research database to the Internet, like Marshfield Clinic's Personalized Medicine Research Project, may provide an added layer of data protection. Such a strategy may appeal to individuals with data security concerns who would otherwise be unwilling to participate. However, this strategy may impede broad and immediate dissemination of the data for legitimate uses and may be seen as an undesirable data-sharing model by advocates of open science.

Also related to privacy, controlled-access initiatives often employ a review process for data requests to reduce the risk of harm from nefarious data uses. Registered access has been proposed as a mechanism to ensure data privacy and security while making data available for research use, particularly data that are browsed rather than downloaded.[31] We observed a dif-

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

ferent model involving a clickwrap agreement through which data seekers click 'I agree' to the terms of use, with no manual review process to authenticate and authorize data seekers. Instant approval through a clickwrap agreement may expedite access to data. But, the implications of implementing this type of agreement to govern the sharing of genomic and other sensitive health data, particularly whether such an agreement is a contract or pseudo-contract and the enforceability of the terms to ensure data (especially downloaded data) are not misused and are adequately protected, should be further explored.[32]

Regarding data access requirements and users, most of the initiatives we reviewed limit access to "qualified researchers" whose intended data use aligns with the goals of the initiative and consent requirements. However, ambiguity remains about how data access determinations are made, and it is often not clearly communicated on the initiative's website whether researchers from government agencies, academic centers, and/or for-profit companies, or even citizen scientists, are eligible to access the data. While some data access requirements might be satisfied by citizen scientists without too much difficulty (e.g., create user account, send email request, submit application), others, such as obtaining IRB approval or executing a data-sharing agreement, pose significant barriers. As support for citizen science in biomedical research increases, the trade-offs involved in making data resources more available to citizen scientists merit further analysis.

It is also not common practice for initiatives to publicize who is accessing the data and how the data are being used. Publicly posting information on approved research protocols and on published studies that utilize data may help participants understand why certain researchers are granted data access and alleviate concerns about potential misuse. It would also help to establish the extent of the benefits of the research and data collection efforts.

Finally, regarding consent, oversight, and participant engagement, the wide utilization of one-time, broad consent captured in our review is consistent with traditional biomedical research practices. Some ethical analysis of large-scale data sharing finds a need for stronger oversight mechanisms and a greater emphasis on participant engagement, including the integration of participants in decision-making processes concerning data sharing.[33] Dynamic consent, which offers granular data-sharing options that can be modified over time to reflect changes in a person's attitudes and beliefs about data sharing, is an example of a proposed tool to cultivate trust and empower participants.[34] Yet only one of the initiatives reviewed describes using this tool. As to broad consent, some have suggested that the ethical justification is tied to participant engagement efforts in general, and a clearly communicated plan for governance involving participant representatives in particular.[35] The JHS CORC is one model for engaging participants, particularly ethnic minorities, in research plus health promotion. However, engaging participants through efforts such as the JHS CORC may be too resource-intensive for other projects and may not be a reproducible model for all data-sharing efforts. Further work is needed to evaluate what participant engagement strategies are most effective and efficient.

Relying solely on publicly-available information may have limited the details captured, but this approach allowed us to view data-sharing practices from the perspective of the public, the people whose data are needed to sustain an MIC. The use of the Not Mentioned Online option on the data collection instrument reflects the variation in the degree of transparency observed on the websites of the initiatives reviewed. Also, this analysis required a thorough, often laborious, examination of websites to find relevant information and understand the different aspects of data sharing discussed above.

There is currently no legal requirement that data-sharing initiatives follow guidelines on transparency and disclose information on data-sharing practices on their websites, or even maintain a website that is accessible to the public via the Internet. However, we recommend disclosure of key information related to data-sharing practices in formats that are easy for the public to find, access, and comprehend. As indicated by the GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data, comprehensive, understandable, and accessible communication of data-sharing practices is important to nurturing trust from participants.[36] Information on public forums, such as a website, may foster trust and encourage prospective participants to share data and enrolled participants to continue contributing data. To support informed decision-making, we also recommend exploration of the feasibility of creating a regularly updated resource for the public that captures key information about data-sharing initiatives, especially those soliciting data directly from the public. Such a guide would facilitate comparisons and function in a manner similar to Charity Navigator with respect to monetary contributions to philanthropic organizations.[37]

Conclusion

Data sharing is essential to creating an MIC that will lead to meaningful benefits through advances in biomedical research, precision medicine, and public health. Without data from individuals, the available

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

resources will be inadequate for forms of scientific inquiry and clinical decision support that depend on big data. Data-sharing initiatives looking to establish or improve an online presence should refer to guidelines to inform web content development and ensure that communication about data-sharing practices is understandable and comprehensive. Public and participant trust in this enterprise rests on transparency about data-sharing practices in domains such as consent, privacy and security, data access, oversight, and participant engagement. Finally, it is also critical that data-sharing initiatives be trustworthy, endeavoring to address human health in a holistic and equitable manner, investing in measures related to privacy and security, experimenting responsibly with graded access mechanisms (including facilitation of citizen science access to at least some kinds of data), and carefully considering the interrelationship of consent processes, oversight structures, and participant engagement efforts.

Note

Dr. McGuire reports personal fees from Geisinger Research, outside the submitted work.

Acknowledgement

We are grateful to project team members who contributed to the development of the data-collection instrument and data collection efforts. The work presented in this article is funded by the National Institutes of Health (NIH) supported the preparation of this article: R01HG008918 (AMG and RCD). The views expressed in this article are those of the authors and do not necessarily reflect those of the NIH.

References

1. National Institutes of Health, "Genomic Data Sharing Policy" (2014), *available at* <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html> (last visited January 10, 2019); D. B. Taichman, J. Backus, C. Baethge, and H. Bauchner et al., "Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors," *Annals of Internal Medicine* 164, no. 7 (2016): 505–506, *available at* <http://annals.org/article.aspx?doi=10.7326/M15-2928> (last visited January 10, 2019).
2. Sage Bionetworks is an example of an open science advocate.
3. A.G. Villanueva, R. Cook-Deegan, B.A. Koenig, and P.A. Deverka et al., "Characterizing the Biomedical Data-Sharing Landscape," *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 21-30.
4. 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.biomedcentral.com/articles/10.1186/s13073-017-0476-3> (last visited January 10, 2019).
5. M.A. Majumder, P.D. Zuk, and A.L. McGuire, "Medical Information Commons," draft, in *Routledge Handbook of the Study of the Commons* (Rochester, NY: Social Science Research Network, Forthcoming 2018), *available at* <https://papers.ssrn.com/abstract=3131913> (last visited January 10, 2019).
6. See GA4GH Framework for Responsible Sharing of Genomic and Health-Related Data (2014), *available at* <https://thehugojournal.springeropen.com/articles/10.1186/s11568-014-0003-1> (last visited January 10, 2019) and Privacy and Security Policy (2015), *available at* <https://www.ga4gh.org/wp-content/uploads/Privacy-and-Security-Policy.pdf> (last visited January 10, 2019). See also OECD Guidelines on Human Biobanks and Genetic Research Databases (2009), *available at* <http://www.oecd.org/sti/emerging-tech/44054609.pdf> (last visited January 10, 2019).
7. See Deverka, *supra* note 4.
8. B.M. Knoppers, "Framework for Responsible Sharing of Genomic and Health-Related Data," *The HUGO Journal* 8, no. 1 (2014): 3, *available at* <https://doi.org/10.1186/s11568-014-0003-1> (last visited January 10, 2019); National Institutes of Health, "Precision Medicine Initiative: Privacy and Trust Principles," All of Us Research Program, *available at* <https://allofus.nih.gov/about/program-overview/precision-medicine-initiative-privacy-and-trust-principles#precision-medicine-initiative-privacy-and-trust-principles-2> (last visited January 10, 2019); See also *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): 704–710, *available at* <https://doi.org/10.1089/omi.2017.0156> (last visited January 10, 2019); Deverka, *supra* note 4; A.L. McGuire, M.A. Majumder, A.G. Villanueva, and J. Bardill et al., "Importance of Participant-Centricity and Trust for a Sustainable Medical Information Commons," *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 12-20; R. Cook-Deegan and A.L. McGuire, "Moving beyond Bermuda: Sharing Data to Build a Medical Information Commons," *Genome Research* 27, no. 6 (2017): 897–901, *available at* <https://doi.org/10.1101/gr.216911.116> (last visited January 10, 2019).
9. See Villanueva, *supra* note 3.
10. See *id.*
11. For description of these categories see Villanueva, *supra* note 3. Definitions reproduced here: open-access data-sharing initiatives offer data access with no required action, such as creating an account or submitting an application. Controlled-access data-sharing initiatives require an action, such as application submission and approval and/or data use agreement, to grant data access.
12. P.A. Harris, R. Taylor, R. Thielke, and J. Payne et al., "Research Electronic Data Capture (REDCap) — A Metadata-Driven Methodology and Workflow Process for Providing Translational Research Informatics Support," *Journal of Biomedical Informatics* 42, no. 2 (2016): 377-381, *available at* <https://doi.org/10.1016/j.jbi.2008.08.010> (last visited January 10, 2019).
13. The Global Alliance for Genomics and Health, *Data Sharing Lexicon* (March 2016), GA4GH Website, *available at* <https://www.ga4gh.org/docs/ga4ghtoolkit/regulatoryandethics/GA4GH_Data_Sharing_Lexicon_Mar15.pdf> (last visited January 10, 2019), at 6.
14. The 34 initiatives reviewed, in alphabetical order, were: Altruist Database; Alzheimer's Disease Neuroimaging Initiative; AmbryShare; Cancer Genome Characterization Initiative; CF Foundation Patient Registry; Colon Cancer Family Registry; Coronary Artery Risk Development in Young Adults; Critical Path Institute CPAD; Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources; Duchenne Registry; Framingham Heart Study; Genes for Good; Global FKRP Registry; HEROIC Registry; Jackson Heart Study; Kaiser Permanente Research Bank; Metastatic Breast Cancer Project; MSSNG; Multiple Myeloma Research Foundation Researcher Gateway; MyVHL: Patient Natural History Study; Open Humans; openSNP; PAH Biobank; Parkinson's Progression Markers Initiative; Personalized Medicine Research Project Database; Project GENIE; Psychiatric Genomics Consortium; Rare Epilepsy Network; Simons Foundation Autism Research Initiative (SFARI Base); Texas Alzheimer's Research and Care Consortium; Texas Cancer Research Biobank; The Genographic Project DNA Analysis

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press

Repository; The Harvard Personal Genome Project; Wisconsin Registry for Alzheimer's Prevention.

15. Zip code is often used as a geographic scale in studies of health and the built environment. Because zip codes can be used to connect data from different databases to examine environmental attributes, zip codes were captured as part of the built environment variable.
16. Of note, some of the initiatives reviewed enrolled participants and collected data in a clinical setting or at an academic institution and shared data electronically online. A few of the initiatives reviewed enrolled participants through the website, and those that provided consent or terms of use information for collecting and sharing research data on minors indicated that the website was intended for adults or that only the parent or legal guardian may enroll a child under 13. Therefore, Children's Online Privacy Protection Act (COPPA) would not be triggered.
17. AliveAndKickn, HEROIC Registry Website, *available at* <https://aliveandkickn.org/heroic-registry-0> (last visited January 10, 2019).
18. See openSNP, Sign Up Website, *available at* <https://opensnp.org/signup> (last visited January 10, 2019).
19. National Institutes of Health, *Certificates of Confidentiality: Background Information*, Research Involving Human Subjects Website, *available at* <https://humansubjects.nih.gov/coc/background> (last visited January 10, 2019).
20. Marshfield Clinic Research Institute, *Frequently Asked Questions*, Personalized Medicine Research Project Website, *available at* <http://www.marshfieldresearch.org/chg/pmrp/faqs> (last visited January 10, 2019).
21. Framingham Heart Study, Genetic Data, Framingham Heart Study Website, *available at* <https://www.framinghamheartstudy.org/fhs-for-researchers/genetic-data/> (last visited January 10, 2019).
22. See the terms of access and information on data access on the AACR Project GENIE: Data Website (September 21, 2018), *available at* <https://www.aacr.org/RESEARCH/RESEARCH/PAGES/AACR-PROJECT-GENIE-DATA.ASPX> (last visited January 10, 2019).
23. Alzheimer's Disease Neuroimaging Initiative, *Data Sharing and Publication Policy* (January 2016) at 1, *available at* <http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf> (last visited January 10, 2019).
24. See "Data Sharing" section in Simons Collection Researcher Distribution Agreement (May 2017), *available at* <http://simonsfoundation.s3.amazonaws.com/share/Policies_and_forms/2017/FormUpdates_2017_06_08/SFARI-RDA.pdf> (last visited January 10, 2019).
25. Categories of members involved in data-sharing decisions are not mutually exclusive.
26. Jackson Heart Study Community Outreach Center Website, *available at* <http://www.jsums.edu/jsucorc/> (last visited January 10, 2019).
27. For discussion on environmental impact on health, see C. Carlsten, M. Brauer, F. Brinkman, and J. Brook et al., "Genes, The Environment and Personalized Medicine," *EMBO Reports* 15, no. 7 (2014): 736–739, *available at* <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4196974/> (last visited January 10, 2019); C. Jiang, X. Wang, X. Li, and J. Inlora et al., "Dynamic Human Environmental Exposome Revealed by Longitudinal Personal Monitoring," *Cell* 175, no. 1 (2018): 277-291.e31, *available at* <https://www.cell.com/cell/fulltext/S0092-8674%2818%2931121-8> (last visited January 10, 2019).
28. See The Exposome in National Research Council, "Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease" (2011): at 44, *available at* <https://www.nap.edu/catalog/13284/toward-precision-medicine-building-a-knowledge-network-for-biomedical-research> (last visited January 10, 2019).
29. Federal Trade Commission, *The Equifax Data: What to Do* (September 8, 2017), FTC Consumer Information Website, *available at* <https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do> (last visited January 10, 2019); S. Morse, "CMS Responds to Data Breach Affecting 75,000 in Federal ACA Portal, *HEALTHCARE FINANCE* (October 22, 2018), *available at* <https://www.healthcarefinancenews.com/news/cms-responds-data-breach-affecting-75000-federal-aca-portal> (last visited January 10, 2019).
30. L. Vaas, "Hacker-besieged DNA data tucked away under military care," *Naked Security*, December 7, 2018, *available at* <https://nakedsecurity.sophos.com/2018/12/07/hacker-besieged-dna-data-tucked-away-under-military-care/> (last visited March 13, 2019).
31. S.O.M. Dyke, E. Kirby, M. Shabani, A. Thorogood, K. Kato, and B.M. Knoppers, "Registered Access: a 'Triple-A' Approach," *European Journal of Human Genetics* 24, no. 12 (2016): 1676–1680, *available at* <https://www.nature.com/articles/ejhg2016115> (last visited January 10, 2019).
32. There is lively discussion in the literature about non-traditional, digital contracts, including R.B. Kar and M.J. Radin, "Pseudo-Contract & Shared Meaning Analysis," *Harvard Law Review* 132 (forthcoming 2019), *available at* <https://ssrn.com/abstract=3124018> (last visited January 10, 2019).
33. 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.
34. I. Budin-Ljøsne, H.J.A. Teare, J. Kaye, and S. Beck et al., "Dynamic Consent: A Potential Solution To Some of The Challenges of Modern Biomedical Research," *BMC Medical Ethics* 18, no. 4 (2017), *available at* <https://doi.org/10.1186/s12910-016-0162-9> (last visited January 10, 2019).
35. B.A. Koenig, "Have We Asked Too Much of Consent?" *Hastings Center Report* 44, no. 4 (2014): 33–34, *available at* <https://onlinelibrary.wiley.com/doi/abs/10.1002/hast.329> (last visited January 10, 2019); S. Garrett, D. Dohan, and B.A. Koenig, "Linking Broad Consent to Biobank Governance: Support from a Deliberative Public Engagement in California," *American Journal of Bioethics* 15, no. 9 (2015): 56–57, *available at* <https://doi.org/10.1080/15265161.2015.1062177> (last visited January 10, 2019).
36. B.M. Knoppers, "Framework for Responsible Sharing of Genomic and Health-Related Data," *The HUGO Journal* 8, no. 3 (2014), *available at* <https://thehugojournal.springeropen.com/articles/10.1186/s11568-014-0003-1> (last visited January 10, 2019).
37. For a discussion on a proposed "researcher reputation system," see 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 10, 2019).

https://doi.org/10.1177/1073110519840482 Published online by Cambridge University Press