

What is a Medical Information Commons?

Juli M. Bollinger, Peter D. Zuk, Mary A. Majumder, Erika Versalovic, Angela G. Villanueva, Rebecca L. Hsu, Amy L. McGuire, and Robert Cook-Deegan

Data sharing is increasingly recognized as a critical component of efforts to understand genomic variation and advance biomedical research. In 2011, the National Academies of Sciences (NAS), an influential advisor to the U.S. government on issues related to science and technology, released a report entitled “Toward Precision Medicine: building a knowledge network for biomedical research and a new taxonomy of disease.”¹ The NAS report advocated for the creation of an “Information Commons” and a “Knowledge Network” to revolutionize biomedical research and clinical care and ultimately improve health outcomes.²

The NAS report embraces the spirit of open science and broad data sharing adopted by the Human Genome Project through its Bermuda Principles, which put forth a set of principles requiring the rapid release of DNA sequence data in publicly-accessible databases.³ The NAS report advocates for an “Information Commons in which data from large populations of patients becomes broadly available for research use.”⁴ The report describes an Information Commons in which phenotypic data acquired from routine clinical care (including electronic health record data) would be integrated with the results of basic biomedical research to advance scientific understanding of health and disease and to improve clinical care.⁵

While the broad goals of the Information Commons and Knowledge Network are clear, many logistical questions regarding structure, organization, and governance remain unanswered. In particular, the NAS

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. **Peter D. Zuk** is a Research Associate in the Center for Medical Ethics and Health Policy at Baylor College of Medicine and Ph.D. candidate in Philosophy at Rice University (both Houston, TX). **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. **Erika Versalovic** is a Ph.D. student in the philosophy department at the University of Washington and a neuroethics fellow with the Center for Neurotechnology in Seattle, WA. **Angela G. Villanueva, M.P.H.**, is a Research Associate at the Center for Medical Ethics and Health Policy at Baylor College of Medicine. **Rebecca L. Hsu** is a research coordinator with the Center for Medical Ethics and Health Policy at Baylor College of Medicine. **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. **Robert Cook-Deegan, M.D.**, is a Professor in the School for the Future of Innovation in Society at Arizona State University.

report purposefully did not detail design features of an Information Commons, instead deferring to a “creative period of bottom-up research activity through pilot projects of increasing scope and scale...from which best practices would emerge.”⁶

Since the release of the NAS report in 2011, many national and international, public and private initiatives have begun to collect and share or facilitate sharing of data on a large scale for research and clinical use (such as, in the U.S. the NIH’s All of Us Research Program, Data Commons, Genomic Data Commons, and Big Data to Knowledge [BD2K] Initiative, 23andMe’s research program, Project Baseline [launched by Verily Life Sciences, Duke University School of Medicine and Stanford Medicine], and internationally, the Global Alliance for Genomics and Health).⁷ Collectively, these efforts may represent the beginning of a “medical information commons” or MIC, defined as “a networked environment in which diverse health, medical, and genomic data on large populations

How should an MIC be organized? Who can have access to the data in an MIC? What types of data should be included? What, if any role, should the participants whose data populate an MIC play in its governance?

become broadly available for research use and clinical applications.”⁸ As the number of entities collecting and sharing data continues to expand and interconnect, questions surrounding social, ethical, legal, and logistical aspects of an MIC have multiplied. Further, although the NAS report did not discuss the work of Elinor Ostrom and others who have contributed to a growing theoretical and empirical literature focused on commons creation and management, that literature also prompts a range of questions: How should an MIC be organized? Who can have access to the data in an MIC? What types of data should be included? What, if any role, should the participants whose data populate an MIC play in its governance?⁹

To explore these questions, we interviewed expert stakeholders involved in various aspects of data-sharing initiatives from diverse employment sectors (i.e. laboratory, academia, non-government organization, government, technology, and healthcare company). In this paper, we report the opinions of these expert stakeholders as to what constitutes (or what features) define an MIC.

Box 1

Working Definition of a Medical Information Commons:

“Medical information commons are networked environments in which diverse sources of health, medical, and genomic data on large populations become broadly available for research use and clinical applications.”

Methods

Semi-structured interviews were conducted to solicit expert opinion on the creation of an MIC and ethical and legal matters concerning data sharing in the context of MIC formation. Interviews occurred between August 2016 and December 2017. This research received approval from the Institutional Review Board at Baylor College of Medicine.

Sample Selection

Purposeful and snowball sampling were employed to identify individuals of interest to interview. Selection of respondents was made by the project team based on relevancy of the respondent’s area of expertise (e.g., medical genetics, information technology, intellectual property law) and role (e.g., leadership within an organization engaged in data sharing or identification as a thought leader). The

sector of the person’s employment was also considered with an understanding that equal representation across all sectors is unlikely since experts on the ethical and legal data-sharing matters concentrate in certain places of employment. Among the respondents are some members of the project’s advisory committee.

Potential respondents received an emailed invitation to participate in a project interview. The invitation sent to individuals unacquainted with the project included a summary of the project’s aims. Interviewees opted to participate via telephone or videoconference and were offered \$100 for their time. Invitations to participate ceased when the simultaneously occurring analysis indicated data saturation.

Data Collection

Prior to the interview, all participants were emailed a list of general questions they would be asked during the interview, a copy of the verbal consent script, and a report of a project meeting summarizing preliminary discussions on the topics addressed during the interview. During the interview, participants were asked to

define an MIC and respond to a working definition of an MIC (see Box 1), describe their ideal vision for an MIC and identify the biggest “non-technical” (e.g., policy/legal, ethical, cultural) challenges to achieving the ideal, and share their views on the role of the individuals whom the data describe (participants) and data ownership. Probing questions for each of the main questions were prepared in advance and conceived in real-time in response to commentary stated during the interview to solicit further input. This paper reports on how experts define an MIC. Data related to what they see as the biggest non-technical challenge, what they think the role of individuals whose data populate an MIC, and how they view data ownership are all published separately.¹⁰

Individuals expressed verbal consent to participate in an audio-recorded interview. The audio recordings were professionally transcribed and all personal identifiers were removed. The project team reviewed the transcripts for accuracy. To foster an open environment, interviewees were offered the option to answer questions off the record. The off-the-record statements are included in the analysis, but excluded from being used as illustrative quotes in this manuscript. While some interviewees permitted public attribution to their statements, all quotes presented here are anonymous.

Data Analysis

Using a grounded theory approach, two members of the research team independently developed thematic codes based on the interview guide and initial interview transcripts. Several measures were employed to ensure consistency across study coders. First, two coders were responsible for independently coding the first five transcripts in order to develop and refine the codebook. During this initial process, the two coders met frequently to compare coding, discuss and resolve any coding differences, and update the codebook accordingly. Second, once the codebook was finalized, the coding pair coded two additional transcripts independently to ensure consistency was achieved. Finally, once coding consistency was demonstrated, the remaining transcripts were assigned both a primary and secondary coder. Coded transcripts were entered into NVivo 11.¹¹ Text pertaining to the definition and characteristics of MICs was organized and analyzed for recurring themes.

Results

Of the 51 interview invitations sent, 40 individuals accepted (78% response rate). One of the interviews involved an additional unanticipated participant, for a total of 41 respondents. Respondents come from all six sectors (Table 1). As noted above, participants also

Table 1

Sectors Represented by Respondents

	Count (%)
Academia	14 (34)
Non-Governmental Organization	9 (22)
Technology Company	8 (20)
Government	4 (10)
Laboratory	3 (7)
Healthcare System	3 (7)

have diverse areas of expertise and role responsibilities. Finally, although our focus was on the U.S., we did include two expert stakeholders from Canada (representing NGO and academic sectors, respectively). Interviews lasted, on average, between 50 and 60 minutes.

Describing a Medical Information Commons

The results presented here correspond to an analysis of responses interviewees offered to the following three questions: 1) “How do you define a medical information commons?” 2) “Is this [working definition, displayed in Box 1, and read to interviewees] the right definition? What are we missing and how would

Table 2

Descriptions of a Medical Information Commons

Collectively-owned, for the public good.

P10: “I think of a commons of as being something that is publicly owned or collectively-owned among a group and collectively accessed.”

P18: “...I think it's a community good and I don't think anybody owns it... That that should be one of the principals up front.”

Organized around the individual, diverse types of data.

P22: “This is sort of my dream world... I think we could do so much good science and address so many public health issues if we had a commons system in which ultimately we could bring as much data as we have on every individual together into a very comprehensive, very deeply descriptive longitudinal record for that individual, and then that would include not just their traditional health data and data generated during when they participated in research but also lifestyle data and fitness data and grocery store receipts... If we just had everything we know about every individual linked together and then have that same information for everybody in the world. We would know everything about every person for everybody.”

Diverse types of data, for the purpose of facilitating research and improving health.

P11: “An [MIC] would represent an aggregate data store of everything possible relevant to improving health of individuals and populations through both research and direct clinical care.”

you refine it?”, and 3) “What is your vision of an ideal medical information commons?”

When asked to define an MIC, most interviewees struggled to articulate a clear vision of an MIC. However, analysis of the varied responses to the questions listed above revealed several key themes: (1) an MIC was seen as a publicly- or collectively-owned resource, (2) for the public good, (3) organized around the individual, (4) containing diverse types of data, (5) for the purpose of facilitating research and improving health. (Table 2).

To anchor further discussion, we asked interviewees for their opinion of our project’s working definition of an MIC, which was based off attributes detailed in the NAS report (Box 1). We use the term “data commons” to describe constituent data initiatives to avoid confusion with the over-arching data-sharing enterprise.

While most respondents supported the working definition, some offered suggestions for improvement. The suggestions reveal insight on features that are both perceived to be essential elements of an MIC, as well the most ethically and logistically problematic.

The majority suggestions fell into four main categories: (1) expand upon the possible uses of an information commons beyond research and clinical applications (e.g., public health), (2) broaden the range of data types that will be included in the commons beyond health, medical, and genomic data (e.g., environmental, behavioral, social), (3) provide language describing stewardship of the data (i.e., someone/some entity is minding the store), and (4) define “broadly available” to clarify who can access the data and the type of data they are permitted to access (e.g., aggregate/summary data, case level, etc.) (Table 3.)

Intended Use

There was broad consensus that an MIC would be a critical resource for fostering scientific research and improving understanding of health and disease. Most interviewees believed that the knowledge gained through collecting and sharing data in an MIC could ultimately benefit the public by leading to improvements in clinical care. For example, one interviewee said: “Every patient who sits down in front of a doctor benefit(s) from the information and experience, so we can put that into a commons pool and get a better understanding. The level of medicine and understanding of the human condition, illness and health will be vastly improved.” (P1)

However, there was disagreement among the stakeholders interviewed as to whether an MIC or its constituent data commons should be designed for purposes beyond biomedical research, including use in clinical care and policy development. Some favored

Table 3

Suggestions to Improve Project Definition of a Medical Information Commons

Expand possible uses

“When you talk about health data and information, some of the information you might collect around lifestyle or broadly cultural environment I don’t think about medical in that sense. I think about health. I think it’s great that you include health in that, but as you continue to define it, I think making sure that it’s not just medical if your goal really is to think broadly about health. That probably gets then into what my ideal would be.”(P24)

Broaden range of data types

“Maybe a health information commons would be more relevant with a very broad definition of health that would include socio-economic status. Do you have home and what is the condition of your home? What is your income? What is your education level? Because all of these things tie into your health status.” (P35)

Describe stewardship of the data

“We do also need to offer certain protections, security protections, privacy protections to the individuals and maybe it’s just making sure that participants who are giving their data are getting something in return or something. There’s something else there that’s about being good stewards of the data. Somebody has to have responsibility for ensuring that the data is secure. I think that’s part of what a commons is about in the knowledge context.”(P16)

Clarify “broadly available”

“I guess the question of what broadly available means is always a tricky one and again I’m thinking of this more from a legal standpoint because broadly available can mean different things in different contexts. Whether broadly available means that anyone who wants it can get it or anyone who reaches certain conditions can get it. Those are the sorts of questions that I would want to know in terms of further clarifying the definition.”(P2)

designing MICs in an open-ended way, to serve multiple purposes and the needs of many different kinds of users.

A medical information commons to me, would represent an aggregate data store of everything possible relevant to improving the health of individuals and populations through both research and direct clinical care and other forms of care, opposing social determinants of health, addressing behavioral and motivational issues. It’s not just the clinical research and the clinical care, but it’s the broader issue of how do we understand what health is all about. That includes the broader suite of social determinants and behavior determinants as well. It should be a research not just for clinicians and researchers, but also for anywhere from grassroots, commu-

nity-based action to national policy and international policy. (P11)

With regard to use in clinical care in particular, some suggested that an MIC should be designed to serve as a resource for clinical decision-making. Others feared that the regulations and requirements necessary for use as a clinical tool would stymie data collection, sharing, and research. These conflicting opinions are illustrated by the following contrasting quotes:

“If we don’t design commons that are flexible enough to support clinical care uses of the commons, we’re screwing it up.” (P21)

“The value of the commons is to advance research overall, not to diagnose a specific patient.” (P26)

Structure/Organization

With respect to the question of how an MIC would be structured/organized, our interview data revealed four major themes: (1) as noted above, most likely there is not a single, integrated MIC, but many data commons and other data initiatives ideally linked together, (2) there is a preference for a federated vs. centralized structure, (3) intended use drives the structure/organization of an MIC, and (4) the role of individuals whose data populate the commons in the governance of the commons should be expanded.

Interviewees pointed to the fact that there are a growing number of data-sharing initiatives already underway and that these initiatives will likely continue to expand in size and interconnectedness. Furthermore, most interviewees believed that numerous practical, logistical, and legal challenges that impede the flow of data (e.g. consent, privacy, ownership, security, international issues) would make it impossible for data to be held in a central location. Instead, data will likely need to stay in place and be linked: “The federated database is probably the most likely outcome. In fact, it’s a certain outcome and we will end up with some kind of federated system, meaning that there will be a number of independent players that have de facto or de jure control over big blocks of data.” (P15)

While some respondents believed that federated models were cumbersome and that centralized models offered significant advantages in efficiency and overall value, others believed that with progress on interoperability, the flexibility associated with federated models could be an asset.¹²

By definition commons are flexible arrangements, and there will be different structures of

commons, and there isn’t a ‘the’ commons, yet there will be hundreds of data commons existing simultaneously, and they may have rather different structures, and I think one of the key attributes you’d want to have is that all of them at least have an opportunity for inter-operability so that you can merge them into an even bigger commons from time to time. There isn’t a single constitution of a commons that’s the right answer. They’re different. Different commons forming groups will adopt different structures and different rules. (P22)

In sum, although there was support for both centralized and federated models for an MIC, the majority of interviewees believed a federated design was more realistic.

Interviewees reported that there were many ways in which a commons can be designed, but that the structure of the commons would be driven by the intended use. “I think use is really important, and all the attributes of an ideal commons flow from the intended use. Without specifying the use, I’m not sure you can specify an ideal commons arrangement. Everything starts with what you are trying to do and what the intended use and hope for benefits are, and then you always would design the commons arrangement to achieve that use.” (P22) The interviewee went on further to describe data commons as “so use-case-specific and so dependent on the challenges of assembling the data... there’s [no] one ideal data structure.” (P22)

Regardless of preference for a central or federated model, several interviewees described a process in which the individuals and communities for whom the data describe are involved in design of the commons. “It seems to be that a patient-centric approach that relies upon patients or participants...being more active in determining how participating in the creation of the commons and determining how information is used seems a useful way of thinking about things.” (P2) One individual stressed that participant-centricity was critical to the success of an information commons. “The one [model] that really popped more often than anything else was the participant-centric [model], how important it is to be participant-centric. If this is going to get to any kind of scale and if it’s going to honor the value of a medical information commons has to be that it’s good for all of us.” (P6)

Access

Most reflective responses to the concept of an MIC described a broadly available resource or collection of resources. However, differences of opinion regarding intended use, structure, governance, and ownership

were responsible for a wide range of opinions about who should have access and on what terms, and, in most cases, a definition of “broadly available” that included significant limitations. For example, in some conversations, “broadly available” was interpreted as referring to the scientific and clinical communities, not the individuals whom the data describe, for-profit companies, or insurance companies. While there was disagreement among interviewees about who should have access and the breadth of that access, there was consensus among the group that this was one of the most challenging aspects of designing commons. While opinions ranged from open access to credentialed/tiered access, the majority of interviewees describe the need for some parameters around access.

I wouldn't necessarily say a clearinghouse, but it's a place where you can come—Reasonably well qualified researchers, right, that's like so there's who can get access to it. I would not think that advertising companies and that kind of thing having access to it would make it a 'medical information commons.' I think it's one built around either researchers or companies that have a pretty strong interest in sort of life science, health, wellness, being able to essentially look at the data in some way, access it, analyze it to answer questions. (P14)

Privacy, security, and ownership, among other issues, were described as complicating factors:

I'm not against public access to some parts of the information. I think that's a good thing, but I think that for different types of data, different levels I guess in terms of comprehensiveness of the information, there would need to be some different controls in terms of purpose of use. There's some type, again, data security issue gets into the privacy issue. (P27)

Data Characteristics

When describing their vision of an MIC, stakeholders discussed data characteristics along four dimensions: types of data, data collection, data quality/integrity, and data/information.

TYPES OF DATA

There was consensus among interviewees that a broader range of data should be collected on factors known to influence health and disease, which go beyond health, medical, and genomic information. Interviewees described a process by which a wide array of data would flow into an MIC from clinical

and research resources. A range of data types were described, including but not limited to health/medical (from electronic health records, direct patient report, wearable devices such as FitBits, etc.), genomic, proteomic, microbiomic, environmental (diet, exposure, exercise, etc.), epidemiologic/mortality, behavioral, socio-demographic, and healthcare administrative (claims data, etc.). A few offered extremely expansive views of relevant data including shopping data, grocery bills, and global positioning system (GPS) data.

DATA COLLECTION

Interviewees agreed that ideal data collection would be longitudinal and reflect the diversity of populations. Some stressed the importance of including data from medically underserved, underrepresented, and minority groups. There was little enthusiasm for MICs to operate under a public health model where individual participation is mandatory or there is an opt-out consent option rather than opt-in. Interviewees commented that an opt-out model would require a high level of trust that is unlikely to be found in the U.S. population given past incidents of research misconduct (and concerns about discrimination in health-care and other domains), especially among minority/underrepresented populations.

DATA INTEGRITY/QUALITY

The issue of data integrity was frequently cited as critical to the value of the commons, with the need for data standardization and curation arising repeatedly: “One of the most valuable things that data communities can...do...is build the community standards. It's that level of standardization where you're setting the expectations at that curatorial level and expectations around what the data should look like that I think is one of the most valuable things that specific communities do as they form commons.” (P4) Interviewees described the challenge of balancing the burdens associated with collecting and curating standardized, high-quality data with the need for data from a broad range of sources to be collected in the least-restrictive manner: “I mean the key issue for commons are least restrictive terms for input, careful curation of and value added, encouragement of re-contribution from use, and then least restrictive terms possible per use.” (P4) There were particular concerns surrounding the quality of data submitted by the public and disagreement over whether or not the data needed to meet clinical use standards: “I really do think we want people to have an opportunity to make sure the data that's contributed is accurate. I'm always surprised when I go in and somebody reads off a list of medications. Either somebody entered it incorrectly or it's

something old, it's something I never took, it's something that didn't work, I feel like we want to make sure information going in is accurate." (P20)

DATA/INFORMATION

Many described data collection, and subsequent data generation, as an iterative process. Implicit in many of the interviewees' descriptions of a Medical Information Commons was that the data generated using the pooled resource would be contributed back into the commons to improve scientific knowledge and, ultimately, health outcomes.

Trust

The need for a trustworthy system emerged as a critical factor to the success of an MIC. Interviewees identified several features critical to building a trustworthy resource: participant involvement, transparency, access, security, and accountability.¹³ "It has to be a

involved in the design, organization, and governance of a commons would greatly enhance both trust in the system and its ultimate success. "If participants can actually be the drivers who are donating their own data to these systems, I think it gets them a clear role and a clear place, a point of engagement that is actually really powerful and will lead to a more sustainable system." (P16)

Discussion

Our interviewees' attitudes align with the NAS envisioned key elements of 1) wide sharing and broad accessibility of data resources, 2) broad usefulness for research with long-view goal of improving health outcomes, and 3) inclusion of clinical and research data. In thinking about details of implementation, interviewees agreed that research discovery is a core aim of an MIC but disagreed on the question of clinical use. Furthermore, interviewees expressed a range

of views regarding the meaning of broad accessibility and just how widely accessible and shared MIC data ought to be. We now discuss the relationship between these views and other major themes that emerged in the interviews, associated ethical and policy/legal considerations, and insights from the literature concerning commons in general and knowledge commons in particular.

Clinical Use

There was a lack of consensus among our interviewees on the question of whether an MIC should be designed and used for clinical purposes. Some believed that clinical use is an essential component of a mature MIC, while others believed that

the commons should be design to maximize research potential. This will be a key issue for ongoing dialogue among relevant stakeholders. Among the salient considerations will be, on the one hand, the additional complexities involved in clinical use (including burdens and tradeoffs vis-a-vis research discovery), and on the other, the potential for clinical applications to yield benefit for individuals whose data make an MIC possible.

Indeed, this lack of consensus is linked to questions about the allocation of resources and distributive justice. Given that available funds are limited, is the additional upfront investment required to ensure that data are "clinical-grade" (i.e., meet the relevant legal standards for clinical use) warranted given the tradeoffs, such as the need to scale back data resources or shift money away from research projects? Will clini-

Our interviewees' attitudes align with the NAS envisioned key elements of 1) wide sharing and broad accessibility of data resources, 2) broad usefulness for research with long-view goal of improving health outcomes, and 3) inclusion of clinical and research data. In thinking about details of implementation, interviewees agreed that research discovery is a core aim of an MIC but disagreed on the question of clinical use.

trusted system. It has to be a transparent system, and it has to be a system that has oversight, and it has to be secure. Everybody's always on this consent and privacy stuff. It's governance and security, that's what we need, and then you will get trust." (P1) Many thought privacy was best addressed through transparent policies. Some advocated for having a clear plan, including avenues of recourse, in the event of a security breach.

Several interviewees believed that participant involvement would foster support for and success of an MIC. "As we learn things...we may be able to synthesize from this medical information commons, we should be able to push information to participants back. I think it will lead to better engagement of the participants because they will see value for themselves." (P35) Furthermore, some believed that having participants or, at a minimum, their representatives,

cians and patients actually access these kinds of data resources, and will they be successful in using these data resources to improve care and outcomes? On the other hand, a recent NAS report urges researchers to return individual-specific results to participants and to take steps to improve the quality of data to support this activity.¹⁴ Why not view a commitment to designing MICs for clinical use as consistent with this trend in policy? MIC designers must also weigh possible benefits to current patients from a resource tailored to clinical needs against possible benefits to future patients from a resource focused on research.

The pluralistic vision that most interviewees articulated suggests one path forward: experimentation with different approaches, and evaluation of whether clinical-grade resources are used by clinicians and patients and lead to improvements in care and outcomes. Whatever one's antecedent view on the matter, there is some reason for optimism given openness on the part of stakeholders to generating and weighing relevant evidence and considering alternative points of view: the process of resolving design conflicts within a commons can spark substantial governance innovations.¹⁵

MIC as Ecosystem

At present, sharing occurs in complex, interrelated networks of open-access and limited common-property data resources. Consider human genomic data. Sensitive (especially individual-level) data are subject to tiered, credentialed access. Research consortia such as the CHARGE Consortium and the Psychiatric Genomics Consortium have procedures for sharing such data among their members and also deposit their datasets in publicly funded repositories to which qualified non-members can apply for access.¹⁶ Summary-level data, on the other hand, are typically readily available in open-access aggregation databases where they can be repurposed for larger analyses. Combined with the findings from a landscape analysis, the interview results reported here suggest that a mature, overarching MIC (a kind of "meta-commons") will be a layered ecosystem, with credentialed access for particularly sensitive data types and open access for less sensitive data types.¹⁷ In other words, it comprises 1) a collection of many limited common-property regimes controlled in various respects (if not fully owned) by individual researchers and consortia, alongside 2) large data-sharing efforts, including open access datasets, and a range of data-sharing facilitators (such as analysis tools).

Balancing Broad Sharing with Other Considerations

Interviewees were clear that the ideal of broad sharing must be achieved in a way that, among other things,

preserves privacy, protects the interests of underserved minority populations, and assures data security and quality.¹⁸ But the devil is in the details. Restrictions on the data that may be accessed, such as denial of access to case-level data providing more detailed phenotypic information that is useful for tracking longitudinal changes (i.e., disease progression) may limit the utility of an MIC. Likewise, restrictions that affect who may access data, such as requirements of an institutional affiliation and approval from an institutional review board that are likely to exclude citizen scientists, may also limit the utility of an MIC. Given the challenge of crafting rules that strike the right balance between privacy and utility, it is important to consider where responsibility and accountability for these kinds of trade-offs should be placed.

In Ostrom's framework, oversight of a common-pool resource is most effectively handled by those appropriating the resource or by those accountable to those appropriating the resource.¹⁹ For many-layered enterprises such as an MIC, oversight responsibilities are pushed down to the lowest (most local) possible level with appropriate capacities. Stakeholders at that level are in the most favorable position to tailor rules to circumstances. These considerations strongly support internal oversight of stored data by those who exercise effective control of them within an MIC. Successful data stewardship requires good security practices from the ground up, especially in the federated arrangements that are likely to characterize MICs for the foreseeable future. Yet more than this alone is required. An MIC is quite different in character from the standard array of common-pool resources. In an MIC, those utilizing relevant resources have clear legal and ethical obligations to non-users, necessitating significant external oversight as well. Crucially, this includes those whom the data describes, the people we have described as participants in an MIC. An MIC must be accountable to and involve those whose contributions make it possible.²⁰

As stressed by our interviewees, a successful MIC also requires measures to assure high data quality. Here the language of stewardship becomes particularly salient. MIC data will often require not just protection, but curation. One particularly prominent model for scientific data stewardship takes the form of the FAIR principles: **f**indability, **a**ccessibility, **i**nteroperability, and **r**eusability.²¹ These principles are readily applicable in an MIC context. For example, longitudinal data collection requires a method of establishing the sameness of an individual in multiple datasets without compromising de-identification (findability). Data must also be readily available within the constraints previously discussed (accessi-

bility). MIC research often also depends on the normalization of data in disparate formats (interoperability). In addition, research discovery often depends on making data available for secondary use, as is common in genomics (reusability).

Flexibility as Asset

Interviewees identified an important degree of flexibility in the concept of a commons, as noted above. This is consistent with the NAS report's "bottom up" model for development of the Information Commons and for determination of best practices. Further, this emphasis on flexibility is consistent with the literature on effective commons management. In particular, Ostrom and colleagues stress the importance of adaptive governance.²² The emergence of differently modeled commons will provide insight into their comparative benefits and drawbacks, provided that there is transparency about performance and investment in an assessment process.

Trust and Trustworthiness

Participant involvement in MIC governance is likely to be crucial not only for the promotion of public trust, but also for the *justification* of that trust. The involvement of these stakeholders in determining MIC aims and ensuring proper oversight of activities in pursuit of those aims provides an MIC with a democratic character that contributes to its being *worthy* of trust.²³ This is a plausible extension of Ostrom's principle that governance procedures for a common-pool resource should be the result of collective agreement by those affected.²⁴ A federated structure of the kind supported by many of the expert stakeholders we interviewed provides multiple levels of accountability, another promising feature for promoting and justifying public trust — provided it does not give rise to diffusion of responsibility.

Limitations

While our sample is composed of individuals from a broad range of sectors informing data sharing policies and the formation of MICs, many interviewees held leadership roles and may not be directly involved in the submission or retrieval of data. A study soliciting input from direct contributors and end-users of the data would bring a different set of perspectives to bear on the development of MICs. Furthermore, most of the individuals interviewed did not represent patient advocacy groups or the public. The interviews described here form part of a larger study that obtained input from the public on matters related data sharing and MIC and are reported elsewhere.²²

Conclusion

Expert stakeholders' attitudes about MICs align with the NAS vision of an Information Commons in which data from large populations is broadly available for research use. However, differences of opinion regarding clinical use, the meaning of broad accessibility, and how widely MIC data should be shared underscore the need for further research to explore potential policy and technological solutions to these outstanding questions.

Note

Ms. Bollinger, Dr. Majumder, Ms. Versalovic, and Ms. Villanueva report grants from National Human Genome Research, during the conduct of the study. Dr. McGuire reports grants from the NIH, during the conduct of the study, and personal fees from Geisinger Research, outside the submitted work.

References

1. J. Halpern, "The U.S. National Academy of Sciences – In Service to Science and Society" *Proceedings of the National Academy of Sciences of the United States of America* 95, no. 5 (1997): 1606-1608.
2. National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (Washington, DC: The National Academies Press, 2011), doi: 10.17226/13284.
3. E. Marshall, "Bermuda Rules: Community Spirit, With Teeth," *Science* 29 (2001): 1192.
4. NAP report, *supra* note 2, at 4.
5. *Id.*, at 2.
6. *Id.*, at 3.
7. National Institutes of Health, *Home Page*, All of Us Research Program website, available at <<https://allofus.nih.gov/>> (last visited September 11, 2018); National Cancer Institute, *GDC Overview*, Genomic Data Commons Website, available at <<https://gdc.cancer.gov/about-gdc/gdc-overview>> (last visited September 11, 2018). National Institutes of Health, *Big Data to Knowledge* (July 2018), Big Data to Knowledge Initiative website, available at <<https://commonfund.nih.gov/bd2k>> (last visited September 11, 2018), 23andMe Research Participation website, available at <<https://www.23andme.com/research/>> (last visited September 11, 2018); Project Baseline, *The Study*, Project Baseline website, available at <<https://www.projectbaseline.com/study/>> (last visited September 11, 2018); Global Alliance for Genomics and Health, *Home Page*, Global Alliance for Genomics and Health website, available at <<https://www.ga4gh.org/>> (last visited September 11, 2018).
8. 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 8, 2019).
9. See M.A. Majumder, P.D. Zuk, and A.L. McGuire, "Medical Information Commons," in B. Hudson, J. Rosenbloom, and D. Cole, eds., *Routledge Handbook of the Study of the Commons* (Taylor & Francis Group, Forthcoming 2019); T. Dietz, E. Ostrom, and P.C. Stern, "The Struggle to Govern the Commons," *Science* 302, no. 5652 (2003): 1907-1912.
10. 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; 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; A.L. McGuire, J. Roberts, S. Aas, B.J. Evans, “Who Owns the Data in a Medical Information Commons?” *Journal of Law, Medicine & Ethics* 47, no. 1 (2019): 62-69.
11. NVivo qualitative data analysis Software; QSR International Pty Ltd. Version 10, 2012.
 12. Typologies in which commons are arrayed on a spectrum from fully centralized to fully decentralized, with some form of federation as the middle way, are common. For a helpful discussion of these distinctions, see, e.g., Contreras and Reichman, “Sharing by Design: Data and Decentralized Commons,” *Science* 350, no. 6266 (2015): 1312-1314, *available at* <<http://science.sciencemag.org/content/350/6266/1312/tab-pdf>> (last visited January 8, 2019).
 13. See A.L. McGuire, *supra* note 10.
 14. National Academies of Sciences, Engineering, and Medicine. “Returning Individual Research Results to Participants: Guidance for a New Research Paradigm,” (Washington, DC: The National Academies Press, 2018), doi.org/10.17226/25094.
 15. E. Ostrom, *Governing the Commons* (Cambridge, United Kingdom: Cambridge University Press, 2015): at 106-114.
 16. 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; P. F. Sullivan, A. Agrawal, C. M. Bulik, and O.A. Andreassen et al., “Psychiatric Genomics: An Update and an Agenda,” *The American Journal of Psychiatry* 175, no. 1 (2017) 15-27.
 17. See Villanueva, *supra* note 16.
 18. See Villanueva, *supra* note 16.
 19. See Ostrom, *supra* note 15 at 93-94.
 20. See Deverka, *supra* note 8.
 21. M. D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, and B. Mons, “The FAIR Guiding Principles for Scientific Data Management and Stewardship,” *Scientific Data* 3 (2016): 160018.
 22. See Majumder, *supra* note 9.
 23. See A.L. McGuire, *supra* note 10.
 24. See Ostrom, *supra* note 15 at 93.