Health Research with Big Data: Time for Systemic Oversight

Effy Vayena and Alessandro Blasimme

In the epoch of big data, biomedical research and health care are set to undergo a massive transformation. Change is affecting the continuum of health-related activities, from the way health data are captured and processed, to how data processing is translated into medical practice, research, and health policy. Pervasive capturing and analysis of diverse data types fuel new capabilities in health research. As a result, data-driven discoveries are expected to play an increasingly important role in biomedical research. The huge variety of big data opens the door to previously unimaginable applications of data mining for disease prevention, diagnosis, and treatment.[1]

Precision medicine, predicated on the creation of knowledge networks from diverse data and disciplines,[2] is one prominent example of the data-driven approaches in biomedical research and healthcare.[3] Not only are such applications of benefit to individuals, they can also lead to more targeted and effective public health interventions. In this respect, *precision public health* is also emerging as a new way of addressing public health needs through leveraging the potential of big data for precision prevention, digital epidemiology, and digital health surveillance.[4] Moreover, the possibility of capturing health-relevant data from innovative sources, such as wellness apps, social networks, smartphones, and wearable technologies, has given rise to the overarching concept of *digital health,* which goes beyond the creation of health-related knowledge through data analyses to health care delivery though digital means.[5]

As these aspirations continue to grow along with the capabilities that fuel them, it becomes difficult to keep up with the challenges they present. Beyond epistemological or technical issues like the value of evidence generated through analysis of disparate data, or the development of robust data security systems, some of the more recalcitrant issues are ethical in nature. Values such as autonomy, privacy, solidarity, and freedom

Effy Vayena, Ph.D., *is professor of Bioethics in the Department of Health Sciences and Technology, ETH Zurich in Switzerland where she directs the Health Ethics and Policy Lab. She is also a faculty associate at the Berkman Klein Center for Internet & Society, Harvard Law School. She holds a Master's in History and Philosophy of Science from Imperial College, London and a Ph.D. in History of Medicine from the University of Minnesota.* **Alessandro Blasimme, Ph.D.**, *is a senior researcher at ETH Zurich in Switzerland, in the Department of Health Sciences and Technology (Health Ethics and Policy Lab). His work revolves around ethical and policy issues in biotechnology and biomedicine. He graduated in Philosophy and received a Master's in Bioethics from "La Sapienza" University of Rome (Italy) as well as a Ph.D. in Bioethics from the University of Milan, European School of Molecular Medicine (Italy).*

119

together with the regulatory strategies nations deploy to protect them appear to be under severe pressure. Consider the familiar example of informed consent — a legal requirement and an ethical cornerstone in research with human participants, which in the big data environment struggles to offer sufficient protections and to safeguard autonomy and privacy. How can a norm as important as consent maintain its role when data are routinely collected online? Can we still rely on consent mechanisms to ensure sufficient levels of control of personal data? What privacy safeguards can consent offer in the age of big data?

Such questions are not unique to the domain of health. On the contrary, they pervade any domain affected by big data. As the European Commission's *Onlife Initiative* points out, to navigate the increasing complexity of life online we have to rethink our conceptual frameworks, including our ethical and legal ones.[6] In the domain of health research, such reflection on our conceptual frameworks can facilitate regulatory and policy solutions to ensure that digitalization will not crowd out ethical values that matter to our society.

In this paper, we propose such a conceptual shift. First, we suggest that health research is set to take place within an evolving health data ecosystem; second, we situate the challenge of informed consent within this evolving ecosystem; and, third, we argue for a *systemic oversight approach* to compensate for the ethical deficits caused by a weakened informed consent.

I. The Evolving Health Data Ecosystem

We propose to conceptualize health data as an evolving ecosystem consisting of three interconnected parts broadly clustered under three rubrics: data sources, data capabilities, and data stakeholders.[7] Although each part gives rise to specific ethical challenges, their interdependence augments the complexity of each challenge.

Data Sources

Health-relevant data are typically data on individuals that have been obtained through a health-related source for clinical or research purposes. Usually such sources include health services (e.g., electronic health records, prescriptions, diagnostic images, insurance data), health research activities (e.g., clinical trial records, biobanks, registries, genomic databases), and public health activities (e.g., immunization records,

disease surveillance, vital statistics, etc.). These data sources have long been used for research purposes. Other sources of data, not necessarily related to health, can also be used for health research. For example, lifestyle and socioeconomic data (such as consumers' profiles from loyalty cards, educational, and financial data), environmental data (e.g., meteorological records and pollution series), and behavioral and social data (e.g., from wellness apps, social media, wearables, etc.) offer new opportunities for capturing

> In this paper, we propose such a conceptual shift. First, we suggest that health research is set to take place within an evolving health data ecosystem; second, we situate the challenge of informed consent within this evolving ecosystem; and, third, we argue for a *systemic oversight approach* to compensate for the ethical deficits caused by a weakened informed consent.

granular aspects of individual health. Linking these various sources can yield a more holistic understanding of health and disease, which in turn could lead to more effective prevention and treatment options.[8]

Health data are often collected under vague terms regarding their future uses. This has been exacerbated by the addition of non-health-related data sources — including data generated by individuals themselves as they act online, use apps and devices equipped with sensing technologies, or generate social media content. Often such technologies are portrayed as tools of empowerment, but they may also introduce new vulnerabilities for individuals, as users produce data for one purpose (e.g., tweets to connect with followers) while being completely unaware of what their data will be used for (e.g., pharmacovigilance research through analysis of Twitter content).[9] These issues are closely linked to a longstanding problem of the online environment, that is, the blurred distinction between the public and the private sphere.[10]

Data Capabilities

Data collection activities have grown in tandem with analytical, technological, and policy capabilities that enable health-related uses of expanded data sources. As vast amounts of data are captured, stored, curated, and linked, novel analytics disclose patterns and infer-

ences about health. Policies and governance structures are also being developed to allow interaction between dispersed and non-interoperable data sources. Health data access and interoperability are currently on the agenda of several leading policy agencies.[11]

New analytic capabilities such as machine learning and neural networks have also begun to be used in biomedical research and health care. The application of these novel methods, however, introduces potential conflicts with some basic tenets of health research. *Black-box medicine*, for instance, employs opaque algorithms for data analysis that may not meet the transparency standards of evidence-based medicine.[12] Moreover, as these methodologies displace the role of human agents, issues of accountability and liability become more puzzling than ever. Insofar as health research is concerned, the use of automated data analysis methods begs for further ethical consideration — in particular, to guard against the possible revival of genetic determinism and racial categories fueled by the ease of drawing correlations within extensive multi-parametric datasets.

Data Stakeholders
Another crucial part of the ecosystem is the increasing number of stakeholders. Some of these stakeholders have always been involved with health data, for example, the pharmaceutical industry or health care providers. Yet, new ones are emerging, such as the data analytics industry providing sophisticated technical capabilities or social media giants sitting on massive amounts of data. These "insurgents" enter the domain of health bringing corporate cultures that are not necessarily aligned with existing regulations in health research.[13]

New dependencies and constellations of power are emerging. Stakeholders take on leading roles in the health data ecosystem due to their advanced (and often proprietary) technical know-how and their role in setting technical standards. Meanwhile, more traditional stakeholders acquire new roles. For example, citizens and patients use the big data capabilities to access and use their own data for their own purposes — as in the case of the Quantified Self movement, which enables individuals to analyze their own data to make discoveries about themselves.[14] Other potentially more controversial uses include individuals making their genetic and genomic data available for a fee. Some genetic and genomic companies interested in brokering consumers' data to biotech firms are already exploring such business models.[15] In these cases, patients can provide the missing link to connect disparate datasets [16] and to enact bottom-up disruptive innovation in health care.[17] Patients organizing their own clinical trials or monetizing their data represent a substantial departure from the investigator-driven model of research premised on participants' altruistic motivations.

II. The Fate of Consent
Access to data for research purposes, and to human data in particular, should meet certain conditions. Of paramount importance among those conditions is informed consent to the collection, storage, use, and sharing of health-related data. This cornerstone of research ethics enacts respect for research participants and promotes their fundamental ethical interests by disclosing information about a study's purpose, risks, procedures (including procedures for handling personal samples and data) as well as measures to cope with harms resulting from participation. Yet, many scholars have questioned the function and the efficacy of informed consent in today's medical and scientific practices. In particular, the role of consent has been challenged in the online environment where data flow from and toward many directions.[18]

In the field of biomedical research, complex regulatory requirements linked to the development and review of informed consent forms are often perceived as unnecessary bureaucratic burdens.[19] Informed consent in clinical research has been considered "culturally biased, legalistic, ritualistic and unevenly enforced."[20] A recent *New York Times* op-ed tellingly referred to it as a "farce."[21] In this context, some see consent as an empty "ceremony of individual control and choice," for which we had better find new alternatives.[22]

Doubts can be raised as to whether individuals can have meaningful control over escalating volumes of research data collected, for instance, in large-scale genomic datasets used by multiple researchers over long periods of time. Moreover, when consenting to have biological samples or data collected in biobanks, individuals often implicitly agree to future uses of their material under rather unspecific terms. Both in the United States and in Europe, for example, stored data and samples can generally be shared and reused without seeking additional consent provided they are deidentified. The U.S. Common Rule "incentivize[s] the use of specimens and information by allowing their use without the knowledge or consent of the individual once they have been stripped of identifiers."[23] These practices are commonplace in the European Union as well,[24] undermining individual control over both the type of studies that will be undertaken and the entities with which samples will be shared.

To address this issue, the National Institutes of Health (NIH) Genomic Data Sharing Policy expects grantees to seek broad consent for reuse and sharing of participants' data and samples, and looks favorably

at the possibility of using dynamic or tiered consent.[25] The recently promulgated revisions to the Common Rule also promote the use of broad consent[26] as an efficient way of ensuring that data and samples are made as widely available as possible for research.[27] In recent years, however, some have cited the theory and practice of informed consent to warrant for other — more radical — possibilities. For example, some have argued that, when risks are minimal, consent to the further use of samples originally donated for research can simply be presumed.[28] Others, instead, have proposed "blanket consent," that is, consent to any future research on one's donated samples, with no further information on future uses, and with no further oversight.[29] "Open consent," adopted by the Personal Genome Project, takes a transparency-based approach and asks donors to have their genetic and genomic data included in an open-access database, with no privacy guarantees.[30] Similarly, the so-called "Portable Legal Consent" allows individuals to share sensitive data — such as genome sequences — with a selected pool of research institutions.[31]

What these models have in common is that they maximize the availability of samples and data, impose little or no restriction on their use, and offer limited information or control to the individual. It is fair to ask whether unspecific consent is truly *informed* consent, a question that has already generated a variety of responses. Among the most popular of them are that unspecific consent promotes the development of science at the cost of "diluting" the ethical interests of donors and research participants;[32] that the ethical soundness of unspecific consent is anchored on the role that well-functioning review boards play in research oversight;[33] or that the better solution to this challenge should be found in interactive types of consent that promise to allow participants to express and update their preference regarding the reuse and the sharing of their data and samples.[34]

These issues are already thorny in the case of research employing what we called standard sources of health data. Further complications arise, however, when standard sources are combined with expanded sources of data produced outside of conventional clinical or academic settings, and collected through commercial devices — such as smartphones and wearable devices — or even over the Internet. In this case, most of the time, informed consent of the sort we are used to seeing in health research is simply not offered. Moreover, Internet users notoriously pay little or no attention at all to online privacy notices, which detail the conditions that people accept to release their data.

This disconnect is only getting worse with the advent of big data.[35] Informed consent was not originally con-ceived for research relying on this type of data, and there are still no established ethical benchmarks to assess the use of data of this sort in the context of research.[36] Moreover, the collection and analysis of such data types can take place without the intermediation of institutional review boards, as in the case of citizen science projects or self-tracking activities.[37] Questions also arise about the efficacy of research consent when organizations possess large amounts of data that can be used for research — Facebook's experiment on emotional contagion is a prominent example.[38] Nonetheless, a number of ethically relevant elements need to be taken into account. For instance, such data are not initially collected for the purpose of medical research, and data subjects cannot opt out from specific types of data processing (including research); nor are they given information regarding potential risks and benefits of research with their data. Moreover, some of the most distinctive forms of analysis in the era of big biomedical data, such as the use of machine learning algorithms, give rise to accountability issues that can hardly be managed through individual consent alone. Under these circumstances, it is hard to imagine how informed consent for medical research — as it is currently practiced — could be relevant to such novel types of research.

As the health data ecosystem expands and evolves, informed consent may leave important ethical interests of research participants unattended. In particular, unspecific informed consent seems to err on the side of providing limited information about the possible uses of personal information. This undermines participants' capacity to exert control on donated data.[39] Insisting on consent while knowing that it does not satisfy informational requirements can create an illusion of control, which may compromise data subjects' rights as well as public trust. What is more, emphasis on maximizing the reuse of samples and data while at the same time reducing the bureaucratic load on science can turn out to be a disincentive to the communication of clinically beneficial information to research participants.[40]

One way of counteracting the tendency to provide less information concerning future uses of biological samples and personal data is to use informed consent models that make room for more individual choice. These solutions have emerged especially in the field of biobanking and genetic research.[41] Authorization-based models and tiered consent, for instance, allow donors to decide upfront which type of studies they approve of,[42] or even which kind of clinical information they want to receive and how they prefer to be recontacted.[43] Others have proposed a partnership-based approach in the field of genetic biobanks that,

after initial broad consent, relies on continued communication between researchers and participants in order to keep the latter informed about ongoing research, collect more data, and inform them of clinically relevant results.[44] "Dynamic consent," another variant, combines choice and ongoing communication, offering a user-friendly web-based interface that allows participants to receive information and update their preferences over time.[45]

Increased granularity allows data subjects to exert more control over the conditions of exposure of their data. This is to be welcomed, as empirical evidence shows research participants are indeed concerned about not having sufficient control on their data.[46] Nonetheless, it also loads research participants with more direct responsibilities for the management of their data. Such control may lead to excessively self-protective behaviors — preventing data from becoming accessible for research[47] — or to over-exposure of personal data – which may harm data subjects by increasing the likelihood of discriminatory practices by insurers and employers. These risks notwithstanding, the value of more granular control over the exposure of self-regarding data cannot be ignored. Yet, as we explain elsewhere, enhanced granularity is more likely to arise from the synergistic effect of new forms of data governance than from innovative consent models alone.[48]

The prospect of linking dispersed data sets raises two important issues: first, authorizations given for the primary data uses (whether the data are from the standard or expanded sources) are unlikely to cover all potential combinations of future uses; second, given the variety of entities involved in big data health research, in terms of both data sources and data capabilities, it is unlikely that the current oversight models will suffice. The combination of weakened consent and insufficient research oversight is potentially detrimental to health research in an evolving health data ecosystem. This state of affairs can result in direct harms to individuals and groups, and also risks damaging public trust in scientific research.

III. The Case for Systemic Oversight

Despite the interest in more granular consent models, there appears to be a tendency toward the widespread adoption of broad consent as the default model for data-driven health research. This is especially true of large-scale initiatives in precision medicine (for instance in the United States, United Kingdom, and Switzerland). The revised Common Rule, promulgated in January 2017, also introduces, for the first time, specific requirements for broad consent forms.[49] To alleviate the tensions built into current data prac-

tices in health research, we argue, oversight mechanisms should adapt to the ecosystemic features of health data.

We consider oversight activities as constitutive elements of health research governance. We use "governance" as a broad term encompassing a variety of approaches aimed at orienting and monitoring the activities of a group of stakeholders without relying exclusively on binding legal norms or pure market mechanisms.[50] *Governance* refers to activities such as self-regulation; soft law; codes of conduct; review bodies; auditing mechanisms; expert advice; coordination initiatives among public authorities, researchers, and private actors; deliberation; citizens' forums; and public engagement initiatives. Some governance mechanisms aim at channeling the activity of stakeholders in ways that correspond to public expectations or to previously declared principles (including ethical principles). These are what we consider the oversight components of health research governance. In different ways and at different levels, oversight mechanisms contribute substantially to the creation and maintenance of public trust in research initiatives. For the sake of brevity, we term the entirety of such components "the oversight pipeline" of health research.

We propose that the oversight pipeline of data-driven health research should be able to shape practices and activities toward socially desirable outcomes. Given the interconnectedness and mutual dependencies of the health data eco*system*, we propose a *systemic* model of oversight for data-driven health research. The health data ecosystem that we have described in this paper is the result of a complex network of convergent technologies whose boundaries are not easy to set. Yet, converging technologies represent an opportunity for the development of innovative governance approaches.[51] Such an approach should track critical points across the spectrum of data-processing activities and actors and remain attentive to the direction that technological development can take, in order to minimize risks for stakeholders while optimizing the delivery of public benefits.

Before we present the substantive features of a systemic orientation to health research oversight, a few preliminary considerations are in order. Systemic oversight is not limited to any particular phase of the data-processing lifecycle, nor is it undertaken by a specific body charged with oversight responsibilities in health research. Rather, systemic oversight spells out the substantive features that oversight activities should have for health research in an evolving data ecosystem. Such features are intended to create alignment throughout the oversight pipeline — for existing bodies and mechanism, as well as for new ones

that may have to be created in the future. Data-driven medical progress could be hindered by differences in regulatory cultures playing out across the wide spectrum of practices and stakeholders that characterize the ecosystem. Given the difficulty and the potential inefficiency of implementing a unique regulatory framework spanning multiple layers of interdependent data-related activities, common ground is needed for the governance of health research. This is more likely to happen at a higher level of abstraction, namely, with the adoption of a substantive point of reference by all the actors involved in research oversight activities. It is our contention that such common ground could foster coordination, increase trust, and prevent systemic crises for the field due to the misalignment of stakeholders' motives and expectations.

The key substantive features of systemic oversight are: 1) the capacity to cope with the uncertainty that surrounds data collection and data uses through *adaptive* and *flexible* mechanisms; 2) the capacity to address the expanded temporality of data-related activities (from storing to re-analysis) through dynamic *monitoring* and *responsiveness*; and finally, 3) the ability to cope with the relational nature of big biomedical data by means of *reflexivity* and *inclusiveness*.

To account for the *uncertainty* that is typical of data-intensive health research, oversight mechanisms should not be seen as procedures for prospective risk assessment, but rather as *adaptive* instruments that respond to change. Recognizing the limits of anticipatory efforts in a fast-moving data ecosystem and the insufficiency of one-size-fits-all solutions to data governance, *adaptivity* refers to the capacity to devise specific oversight mechanisms for new data types and uses. The *flexibility* requirement, by contrast, demands that, for oversight purposes, data are treated under different rubrics depending on their actual use rather than their source. Adaptive and flexible oversight will enhance effective containment of harmful effects in case of security failures (such as data breaches, forced re-identification, and the like).

Large-scale data collection allows virtually infinite possibilities and combinations of data analysis methods that can be used to search for medically relevant correlations in datasets. The power of such analytical capabilities lies in the fact that they can be invented on the go. While certain parameters are set, it is hard to predict what future analytical developments will look like. This undermines the possibility of explicitly defining in advance the purpose of data collection and the commitment to limit data use to that purpose,[52] thus endangering the creation of public trust around data-driven health research. It is also difficult to specify in advance whether data protection mea-

sures taken to ensure the privacy of data subjects can indeed achieve this goal. The effectiveness of privacy impact assessments (PIA), a well-established practice to ascertain privacy risks in data projects, has been challenged for this very reason.[53] Moreover, the opacity of algorithmic approaches, black-box medicine and the like, indicates lurking risks and the need for better accountability mechanisms.[54] Technological transformations are faster paced than in the previous world of standard data sources and analytics. At the same time, it is necessary, especially in the scenario we are depicting, to acknowledge the limits of anticipatory practices.[55] Since important technological players are bound to have sufficient resources to shape the field to their advantage, regulators may have a hard time trying to fix the downstream effects of such activities. Greater coordination is thus required between public authorities, academic players, and private actors in order to set up converging oversight practices able to adapt rapidly to evolving technological conditions.

Ethical oversight of data collection and use should be a continuous *monitoring* process aimed at detecting signals of novel vulnerabilities for research participants, such as potential harms of discrimination and new privacy risks. Ethical challenges exist throughout the whole data-processing lifecycle. Moreover, each instance of data analysis gives birth to more data with renewed life expectancy. These developments can instigate ethical transgressions that cannot be anticipated at the moment of authorized collection. *Responsive* measures should aim at preventing vulnerabilities from resulting in harm to individual participants or their community, as well as at containing and compensating harms should they occur. Such measures, however, cannot be entirely established during the evaluation phase of a given research project. Rather they need to be revised regularly to align with the expanded temporality of multiple data reuse.

A fundamental tenet of big data research is the identification of patterns and correlations within large collections of structured and unstructured data. In essence, such patterns will disclose not only how phenomena relate to one another but also how individuals relate to others, or groups of individuals to other groups of individuals. A typical biomedical big data project explores correlations among phenotypic, biological, and behavioral data to ascertain disease patterns. Revealing such relations can quickly implicate individuals and groups of people who were never aware of underlying correlations linking them to one another. This may have especially sensitive implications for both individuals and groups in biographical, cultural, ethnic, and racial terms. Nevertheless, it is impossible for the current one-stop shop model of

ethical assessment to anticipate those outcomes and include all these features in the assessment of foreseeable risks.

Oversight activites should be *reflexive* enough to consider the effects of classificatory practices as they emerge from scientific activities (including their historical load and cultural consequences). It is crucial to cultivate awareness of how and why previous governance models have failed in this respect or have ended up reinforcing dominant interests and power structures over individual entitlements, communities' interests, and the common good.[56] Reflexive oversight should thus foster the critical appraisal of the assumptions underlying previous regulatory arrangements.[57] Reflexivity can be fostered by engaging relevant stakeholders upstream and by *including* lay or traditionally under-represented voices (such as the voice of patients or that of socially marginalized communities) to promote mutual learning before technological path dependencies[58] become established.

IV. Points to Consider

Our systemic oversight model is not an isolated example of an approach providing a common ground for a variety of governance actors and activities. To begin with, the contribution of governance to social coordination, accountability, and public trust has long been established in political science and public policy studies. The so-called "new governance" (introduced in the United Kingdom in the 1990s) relies on public engagement of civil society actors, integration across government levels and between government and private actors, decentralization, deliberation, use of soft law instruments, experimentation, and mutual learning.[59]

Many of these features are of relevance to systemic oversight as well. Interestingly, new approaches to governance have enjoyed widespread consideration in the field of science policy, in particular with respect to the public management of environmental issues. Approaches such as public deliberation[60] and dynamic oversight[61] in nanobiotechnology, reflexive governance for ecological problems,[62] and constructive technology assessment[63] all stress the "limits of prognostic knowledge"[64] and the importance of continued learning to redress power asymmetries among stakeholders. Moreover, some substantive features of systemic oversight like reflexivity, inclusivity, and responsiveness, are also characteristic of the Responsible Research and Innovation framework,[65] which

aimed at promoting the democratic governance of science and innovation.

Innovative forms of governance have also been discussed in the domain of health research. "Adaptive governance," for instance, conceptualizes research participants as a collective body to be consulted and included in the definition of genomic biobank policies.[66] One frequently proposed way to keep research governance structures in line with technological, social, and regulatory change is to adopt deliberative democracy practices. In biobank research, it has been proposed that consent should not be seen as a mechanism for individual choice, but rather as "consent to a governance scheme."[67] Deliberative democracy techniques — like citizens' forums — are tools that allow lay publics to provide input. Such forums can also become institu-

> Oversight activites should be *reflexive* enough to consider the effects of classificatory practices as they emerge from scientific activities (including their historical load and cultural consequences). It is crucial to cultivate awareness of how and why previous governance models have failed in this respect or have ended up reinforcing dominant interests and power structures over individual entitlements, communities' interests, and the common good.

tionalized and take the form of "citizen-led Community Advisory Boards" endowed with the right to propose policies on topics such as secondary findings, communication with participants, and the like.[68] The collective engagement of research participants in decisions about data governance (including the creation of guiding ethical and privacy principles) seems particularly well suited to the assemblage of large-scale repositories – such as those envisaged in flagship programs like the "All of Us" Precision Medicine Initiative.[69]

In addition, in the field of health research governance, data access review has been proposed as a mechanism to assess precisely the risks connected with the re-use of personal data. Review processes can be focused on assessing coding and security measures[70] or can instead embrace a more comprehensive evaluation as in the case of Data Access Committees (DACs) in genetic and genomic biobanking. DACs typically evaluate the fit between data access requests

and the original purpose of collection, make sure that those seeking access are technically qualified, stipulate data access agreements, and monitor possible violations.[71] Along similar lines, Grady and colleagues endorse broad consent upon condition that future uses are overseen before approval (ongoing oversight) and that, whenever possible, communication with participants, and enforcement and monitoring activities take place.[72]

These models of expanded and more effective governance in health research are laudable. Many of them possess some of the substantive features of systemic oversight. Yet, they have been proposed for standard sources of data, and they focus either on policy-setting mechanisms for large-scale research initiatives or on review procedures for the use of research resources.

Nevertheless, the required transformations are no greater than the scientific, medical and societal revolutions that big data are causing globally. The scale of the digital revolution in medicine and medical research is such that remodeling regulatory standards with the aim of steering innovation is not only desirable but rather imperative.

The legal landscape around the use of personal data — including for research purposes — is evolving and will continue to evolve in the years to come, as evidenced by the European General Data Protection Regulation (GDPR)[73] — adopted in 2016 and to become operative in 2018 — or by the planned revision of the Swiss law on data protection.[74] New legislation may also introduce new elements to the existing oversight pipeline, as is the case with the GDPR demanding the

> The complexity of the health data ecosystem will require concerted effort to address emerging ethical issues. Enhancing the granularity of consent should definitely continue to be pursued. In addition, innovative research governance models need to be given appropriate consideration and tested in practice. However, the new reality of big data research within the evolving data ecosystem demands ethical controls that are spread throughout the continuum of regulatory activities and can respond to unexpected events across the life cycle of data use. We have argued that this can be achieved through the implementation of a systemic oversight approach tailored to the features of the health data ecosystem.

Systemic oversight instead is intended for issues arising when data sources are mixed or when big data research is conducted by organizations that are not conventional players in the field of health research. Moreover, systemic oversight cuts across policy-making and operative decisions at the level of local research initiatives. Finally, this approach considers all data types as potentially sensitive data — irrespective of the context from which they were sourced.

Implementing the principles of systemic oversight across the spectrum of regulatory activities taking place in data-driven health research will certainly have numerous procedural implications, which we do not discuss here. Suffice it to say that oversight bodies adopting a systemic orientation will need to apply the principles of systemic oversight to their composition, their mandate, their operational mechanisms, and their enforcement and monitoring powers. Undoubtedly, this process of adaptation and substantive convergence toward systemic oversight will be complex.

presence of data protection officers (DPOs) at each research institution.[75] While those developments are necessary, legal solutions alone cannot address the full complexity of the heterogeneous ecosystem of data-driven research. As noted by Laurie, legal frameworks for specific kinds of health practices (e.g., biobanking) have often proved ineffective and burdensome.[76] Moreover, nation-specific legal approaches to health data management may hinder cross-border collaboration and exchange. New legally required oversight components — like DPOs — will need to be aligned with the other pieces of the pipeline.

Efficient governance will require a common orientation between the various components of the oversight pipeline.[77] Systemic oversight endorses the idea of distributed responsibility for the processing of health data. This may lead to difficulty in tracing responsibility for decisions that affect different stakeholders in different ways. To avoid the dilution of accountability, we argued that the systemic oversight processes

should remain modifiable through the input of concerned stakeholders (inclusivity).

The complexity of the health data ecosystem will require concerted effort to address emerging ethical issues. Enhancing the granularity of consent should definitely continue to be pursued. In addition, innovative research governance models need to be given appropriate consideration and tested in practice. However, the new reality of big data research within the evolving data ecosystem demands ethical controls that are spread throughout the continuum of regulatory activities and can respond to unexpected events across the life cycle of data use. We have argued that this can be achieved through the implementation of a systemic oversight approach tailored to the features of the health data ecosystem. Systemic oversight, we have argued, can become a catalyst for ethical health research in the big data era and holds the key to public trust in research. We have sketched the substantive cornerstones of such an approach without discussing the necessary procedural adaptations. Our aim is to generate discussion around the conceptual shift toward the systemic oversight approach, to make progress toward its operationalization.

Note

Acknowledgments

References
1. F. S. Collins and H. Varmus, "A New Initiative on Precision Medicine," *New England Journal of Medicine* 372, no. 9 (2015): 793–795; E. Elenko, L. Underwood, and D. Zohar, "Defining Digital Medicine," *Nature Biotechnology* 33, no. 5 (2015): 456–461; S. Hawgood, I. G. Hook-Barnard, T. C. O'Brien, and K. R. Yamamoto, "Precision Medicine: Beyond the Inflection Point," *Science Translational Medicine* 7, no. 300 (2015): 300ps17; A. Blasimme and E. Vayena, "'Tailored-to-You': Public Engagement and the Political Legitimation of Precision Medicine," *Perspectives in Biology and Medicine* 59, no. 2 (2017): 172–188.
2. National Research Council, *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease* (Washington, D.C.: National Academies Press, 2011).
3. Blasimme and Vayena, *supra* note 1; A. Blasimme and E. Vayena, "Becoming Partners, Retaining Autonomy: Ethical Considerations on the Development of Precision Medicine," *BMC Medical Ethics* 17, no. 1 (2016): 67.
4. J. S. Brownstein, C. C. Freifeld, and L. C. Madoff, "Digital Disease Detection — Harnessing the Web for Public Health Surveillance," *New England Journal of Medicine* 360, no. 21 (2009): 2153–2157; M. Salathé et al., "Digital Epidemiology," *PLOS Computational Biology* 8, no. 7 (2012): e1002616; E. Vayena, M. Salathé, L. C. Madoff, J. S. Brownstein, "Ethical Challenges of Big Data in Public Health," *PLoS Computational Biology* 11, no. 2 (2015): e1003904, doi:10.1371/journal.pcbi.1003904; M. J. Khoury, M. F. Iademarco, and W. T. Riley, "Precision Public Health for the Era of Precision Medicine," *American Journal of Preventive Medicine* 50, no. 3 (2016): 398–401; A. Flahault, A. Geissbuhler, I. Guessous, P. Guérin, I. Bolon, M. Salathé, and G. Escher, "Precision Global Health in the Digital Age," *Swiss Medical Weekly* 147 (2017): w14423.
5. A. Blasimme, T. Haüserman, A. Adjekum, and E. Vayena, "Digital Health: Meeting the Ethical and Policy Challenges," *Swiss Medical Weekly* (in press); E. Underwood and D. Zohar, "Defining Digital Medicine," *Nature Biotechnology* 33, no. 5 (2015): 456–461.
6. L. Floridi, ed., *The Onlife Manifesto: Being Human in a Hyperconnected Era* (New York: Springer, 2014): at 7.
7. World Health Organization, The Health Data Ecosystem and Big Data, *available at* <http://www.who.int/ehealth/resources/ecosystem/en/> (last visited January 24, 2018); E. Vayena, J. Dzenowagis, J.S. Brownstein, and A. Sheikh, "Policy Implications of Big Data in the Health Sector," *Bulletin of the World Health Organization* 96, no. 1 (2018): 66-68.
8. S. H. Jain, B. W. Powers, J. B. Hawkins, and J. S. Brownstein, "The Digital Phenotype," *Nature Biotechnology* 33, no. 5 (2015): 462–463; E. Vayena and U. Gasser, "Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in Biomedicine," in B. D. Mittelstadt and L. Floridi, eds., *The Ethics of Biomedical Big Data* (Cham, Switzerland: Springer International Publishing, 2016): 17-39.
9. C. C. Freifeld, J. S. Brownstein, C. M. Menone, W. Bao, R. Filice, and T. Kass-Hout, "Digital Drug Safety Surveillance: Monitoring Pharmaceutical Products in Twitter," *Drug Safety* 37, no. 5 (2014): 343–350.
10. D. O'Brien, J. Ulman, M. Altman, U. Gasser, M. Bar-Sinai, K. Nissim, S. Vadhan, M. J. Wocik, and A. Wood, "Integrating Approaches to Privacy Across the Research Lifecycle: When Is Information Purely Public?" Berkman Center Research Publication, Paper no. 2015-7 (2015), *available at* <https://papers.ssrn.com/abstract=2586158> (last visited December 27, 2017).
11. Organisation for Economic Co-operation and Development, "Recommendation on Health Data Governance," (2017), *available at* <http://www.oecd.org/els/health-systems/health-data-governance.htm> (last visited January 24, 2018).
12. I. G. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo, "The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care," *Health Affairs* 33, no. 7 (2014): 1139–1147; W. Nicholson Price II, "Black-Box Medicine," *Harvard Journal of Law & Technology* 28, no. 2 (2015): 419-467.
13. Anonymous, "A Digital Revolution in Health Care Is Speeding up," *The Economist*, March 2, 2017, *available at* <https://www.economist.com/news/business/21717990-telemedicine-predictive-diagnostics-wearable-sensors-and-host-new-apps-will-transform-how> (last visited January 24, 2018).
14. D. Nafus, ed., *Quantified: Biosensing Technologies in Everyday Life* (Cambridge, MA: MIT Press, 2016): at ix.
15. J. L. Roberts, S. Pereira, and A. L. McGuire, "Should You Profit from Your Genome?" *Nature Biotechnology* 35, no. 1 (2017): 18–20.

16. G. M. Weber, K. D. Mandl, and I. S. Kohane, "Finding the Missing Link for Big Biomedical Data," *JAMA* 311, no. 24 (2014): 2479–2480.

17. E. Hafen, "Midata Cooperatives — Citizen-Controlled Use of Health Data Is a Prerequisite for Big Data Analysis, Economic Success and a Democratization of the Personal Data Economy," Abstract presented at 9th European Congress on Tropical Medicine and International Health, September 6-10, 2015, Basel, Switzerland, in *Tropical Medicine & International Health* 20, Suppl. 1 (2015): 129.

18. E. Vayena, A. Mastroianni, and J. Kahn, "Caught in the Web: Informed Consent for Online Health Research," *Science Translational Medicine* 5, no. 173 (2013): 173fs6.

19. E. C. Hayden, "A Broken Contract," *Nature* 486, no. 7403 (2012): 312–314.

20. G. E. Henderson, "Is Informed Consent Broken?" *American Journal of the Medical Sciences* 342, no. 4 (2011): 267–272.

21. M. A. Sekeres and T. D. Gilligan, "Informed Patient? Don't Bet On It," *New York Times*, March 1, 2017, *available at* <https://www.nytimes.com/2017/03/01/well/live/informed-patient-dont-bet-on-it.html?_r=0> (last visited January 24, 2018).

22. B. A. Koenig, "Have We Asked Too Much of Consent?" *Hastings Center Report* 44, no. 4 (2014): 33-34.

23. The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine, *available at* <https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf> at 81 (last visited January 24, 2018).

24. B. S. Elger and A. L. Caplan, "Consent and Anonymization in Research Involving Biobanks: Differing Terms and Norms Present Serious Barriers to an International Framework," *EMBO Reports* 7, no. 7 (2006): 661-666.

25. National Institutes of Health, Genomic Data Sharing Policy, NOT-OD-14-124 (August, 27, 2014), *available at* <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html> (last visited December 27, 2017).

26. A. Blasimme, C. Moret, S. A. Hurst, and E. Vayena, "Informed Consent and the Disclosure of Clinical Results to Research Participants," *American Journal of Bioethics* 17, no. 7 (2017): 58-60.

27. C. Grady, L. Eckstein, B. Berkman, D. Brock, R. Cook-Deegan, S. M. Fullerton, H. Greely, M. G. Hansson, S. Hull, B. Lo, R. Pentz, L. Rodriguez, C. Weil, B. S. Wilfond, and D. Wendler, "Broad Consent for Research With Biological Samples: Workshop Conclusions," *American Journal of Bioethics* 15, no. 9 (2015): 34–42.

28. D. Wendler and E. Emanuel, "The Debate over Research on Stored Biological Samples: What Do Sources Think?" *Archives of Internal Medicine* 162, no. 13 (2002): 1457–1462.

29. D. Wendler, "Broad versus Blanket Consent for Research with Human Biological Samples," *Hastings Center Report* 43, no. 5 (2013): 3-4; T. Tomlinson, "Respecting Donors to Biobank Research," *Hastings Center Report* 43, no. 1 (2013): 41–47.

30. J. E. Lunshof, R. Chadwick, D. B. Vorhaus, and G. M. Church, "From Genetic Privacy to Open Consent," *Nature Reviews Genetics* 9, no. 5 (2008): 406–411.

31. See Vayena et al., *supra* note 18.

32. V. Árnason, "Coding and Consent: Moral Challenges of the Database Project in Iceland," *Bioethics* 18, no. 1 (2004): 27–49; B. Hofmann, "Broadening Consent — and Diluting Ethics?," *Journal of Medical Ethics* 35, no. 2 (2009): 125–129; J. R. Karlsen, J. H. Solbakk, and S. Holm, "Ethical Endgames: Broad Consent for Narrow Interests; Open Consent for Closed Minds," *Cambridge Quarterly of Healthcare Ethics* 20, no. 4 (2011): 572–583.

33. G. Helgesson, "In Defense of Broad Consent," *Cambridge Quarterly of Healthcare Ethics* 21, no. 1 (2012): 40–50.

34. Food and Drugs Administration, "Use of Electronic Informed Consent: Questions and Answers. Guidance for Institutional Review Boards, Investigators and Sponsors" *available at* <https://www.fda.gov/downloads/drugs/guidances/

ucm436811.pdf> (last visited January 24, 2018); J. Kaye, E. A. Whitley, D. Lund, M. Morrison, H. Teare, and K. Melham, "Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks," *European Journal of Human Genetics* 23, no. 2 (2015): 141-146.

35. F. H. Cate and V. Mayer-Schönberger, "Notice and Consent in a World of Big Data," *International Data Privacy Law* 3, no. 2 (2013): 67–73.

36. See Vayena et al., *supra* note 18.

37. See Nafus, *supra* note 14.

38. J. P. Kahn, E. Vayena, and A. C. Mastroianni, "Opinion: Learning as We Go: Lessons from the Publication of Facebook's Social-Computing Research," *Proceedings of the National Academy of Sciences* 111, no. 38 (2014): 13677–13679.

39. E. Vayena and A. Blasimme, "Biomedical Big Data: New Models of Control over Access, Use and Governance," *Journal of Bioethical Inquiry* 14, no. 4 (2017): 1-13.

40. Blasimme et al., *supra* note 26; A. Blassime, A. Soulier, S. Julia, S. Leonard, and A. Cambon-Thomsen, "Disclosing Results to Genomic Research Participants: Differences That Matter," *American Journal of Bioethics* 12, no. 10 (2012): 20-22; E. Vayena and J. Tasioulas, "Genetic Incidental Findings: Autonomy Regained?" *Genetics in Medicine* 15, no. 11 (2013): 868-870.

41. V. Anastasova, A. Blasimme, S. Julia, and A. Cambon-Thomsen, "Genomic Incidental Findings: Reducing the Burden to Be Fair," *American Journal of Bioethics* 13, no. 2 (2013): 52-54.

42. T. Caulfield, R. E. G. Upshur, and A. Daar, "DNA Databanks and Consent: A Suggested Policy Option Involving an Authorization Model," *BMC Medical Ethics* 4, no. 1 (2003): E1.

43. E. M. Bunnik, A. Cecile, J. W. Janssens, and M. H. N. Schermer, "A Tiered-Layered-Staged Model for Informed Consent in Personal Genome Testing," *European Journal of Human Genetics* 21, no. 6 (2013): 596–601.

44. A. L. McGuire and L. M. Beskow, "Informed Consent in Genomics and Genetic Research," *Annual Review of Genomics and Human Genetics* 11 (2010): 361–381.

45. J. Kaye, E. A. Whitley, D. Lund, M. Morrison, H. Teare, and K. Melham, "Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks," *European Journal of Human Genetics* 23, no. 2 (2015): 141–146; I. Budin-Ljøsne, H. J. A. Teare, J. Kaye, S. Beck, H. B. Bentzen, L. Caenazzo, C. Collett, F. D. Abramo, H. Felzmann, T. Finlay, M. K. Javaid, E. Jones, V. Katic, A. Simpson, and D. Mascalzoni, "Dynamic Consent: A Potential Solution to Some of the Challenges of Modern Biomedical Research," *BMC Medical Ethics* 18, no. 1 (2017): 4

46. See Kahn et al., *supra* note 38.

47. K. H. Jones, G. Laurie, L. Stevens, C. Dobbs, D. V. Ford, and N. Lea, "The Other Side of the Coin: Harm due to the Non-Use of Health-Related Data," *International Journal of Medical Informatics* 97, no. 1 (2017): 43–51.

48. See Kahn et al., *supra* note 38.

49. J. Sugarman, "Examining Provisions Related to Consent in the Revised Common Rule," *American Journal of Bioethics* 17, no. 7 (2017): 22–26.

50. "Governance" has been defined by the Club of Rome as "the command mechanism of a social system and its actions that endeavor to provide security, prosperity, coherence, order and continuity to the system…. Taken broadly, the concept of governance should not be restricted to the national and international systems but should be used in relation to regional, provincial and local governments as well as to other social systems." A. King and B. Schneider, *The First Global Revolution. A Report of the Council of the Club of Rome* (New York: Pantheon Books, 1991): at 114. Other definitions aptly tone down the hierarchical and directive sense conveyed by the expression "command mechanisms" and emphasize — as we also propose — the centrality of steering and control mechanisms to the idea of governance. See, e.g., J. N. Rosenau,

"Governance in the Twenty-First Century," *Global Governance* 1, no. 1 (1995): 13–43.

51. P. Macnaghten, M. B. Kearnes, and B. Wynne, "Nanotechnology, Governance, and Public Deliberation: What Role for the Social Sciences?" *Science Communication* 27, no. 2 (2005): 268–291.

52. G. Laurie, "Reflexive Governance in Biobanking: On the Value of Policy Led Approaches and the Need to Recognise the Limits of Law," *Human Genetics* 130, no. 3 (2011): 347-356.

53. J. Polonetsky, O. Tene, and J. Jerome, "Benefit-Risk Analysis for Big Data Projects," *Future of Privacy Forum* (September, 2014), *available at* <https://fpf.org/wp-content/uploads/FPF_DataBenefitAnalysis_FINAL.pdf> (last visited August 18, 2017).

54. F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge: Harvard University Press, 2015): at 51; Price, *supra* note 12.

55. P. M. Macnaghten, M. B. Kearnes, and B. Wynne, "Nanotechnology, Governance, and Public Deliberation," *Science Communication* 27, no. 2 (2005): 268-291.

56. J.-P. Voß and B. Bornemann, "The Politics of Reflexive Governance: Challenges for Designing Adaptive Management and Transition Management," *Ecology and Society* 16, no. 2 (2011): 9.

57. C. M. Hendriks and J. Grin, "Contextualizing Reflexive Governance: The Politics of Dutch Transitions to Sustainability," *Journal of Environmental Policy & Planning* 9, no. 3 (2007): 333–350.

58. D. Collingridge, *The Social Control of Technology* (New York: St. Martin's Press, 1980): at 47.

59. R. A. W. Rhodes, "The New Governance: Governing without Government," *Political Studies* 44, no. 4 (1996): 652–667; J. Scott and D. M. Trubek, "Mind the Gap: Law and New Approaches to Governance in the European Union," *European Law Journal* 8, no. 1 (2002): 1–18.

60. P. Macnaghten, M. B. Kearnes, and B. Wynne, "Nanotechnology, Governance and Public Deliberation: What Role for the Social Sciences?" *Science Communication* 27, no. 2 (2005): 268-291.

61. G. Ramachandran, S. M. Wolf, J. Paradise, J. Kuzma, R. Hall, E. Kokkoli, and L. Fatehi, "Recommendations for Oversight of Nanobiotechnology: Dynamic Oversight for Complex and Convergent Technology," *Journal of Nanoparticle Research* 13, no. 4 (2011): 1345–1371.

62. Voß and Bornemann, *supra* note 56.

63. A. Rip, J. Schot, and T. J. Misa, "Constructive Technology Assessment: A New Paradigm for Managing Technology in Society", in A. Rip, J. Schot, and T. J. Misa (eds.), *Managing Technology in Society. The Approach of Constructive Technology Assessment* (New York: Pinter Publishers, 1995): 1-14; D. H. Guston and D. Sarewitz, "Real-Time Technology Assessment," *Technology in Society* 24, nos. 1–2 (2002): 93–109; J. Grin and H. van de Graaf, "Technology Assessment as Learning," *Science, Technology & Human Values* 21, no. 1 (1996): 72–99; A. Genus, "Rethinking Constructive Technology

Assessment as Democratic, Reflective, Discourse," *Technological Forecasting and Social Change* 73, no. 1 (2006): 13–26.

64. Voß and Bornemann, *supra* note 56.

65. J. Stilgoe, R. Owen, and P. Macnaghten, "Developing a Framework for Responsible Innovation," *Research Policy* 42, no. 9 (2013): 1568–1580.

66. K. C. O'Doherty, M. M. Burgess, K. Edwards, R. P. Gallagher, A. K. Hawkins, J. Kaye, V. McCaffrey, and D. E. Winckoff, "From Consent to Institutions: Designing Adaptive Governance for Genomic Biobanks," *Social Science & Medicine* 73, no. 3 (2011): 367–374.

67. See Sekeres and Gilligan, *supra* note 21.

68. S. B. 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; E. Vayena, R. Brownsword, S. J. Edwards, B. Greshake, J. P. Kahn, N. Ladher, J. Montgomery, D. O'Connor, M. P. Richards, A. Rid, M. Sheehan, P Wicks, and J Tasioulas, "Research Led by Participants: A New Social Contract for a New Kind of Research," *Journal of Medical Ethics* 42, no. 4 (2016): 216–219.

69. B. J. Evans, "Barbarians at the Gate: Consumer-Driven Health Data Commons and the Transformation of Citizen Science," *American Journal of Law & Medicine* 42, no. 4 (2016): 651-658; B. J. Evans, "Power to the People: Data Citizens in the Age of Precision Medicine," *Vanderbilt Journal of Entertainment & Technology Law* 19, no. 2 (2017): 243-265.

70. M. G. Hansson, J. Dillner, C. R. Bartram, J. A. Carlson, and G. Helgesson, "Should Donors Be Allowed to Give Broad Consent to Future Biobank Research?" *The Lancet Oncology* 7, no. 3 (2006): 266–269.

71. M. Shabani, B. M. Knoppers, and P. Borry, "From the Principles of Genomic Data Sharing to the Practices of Data Access Committees," *EMBO Molecular Medicine* 7, no. 5 (2015): 507–509.

72. C. Grady, L. Eckstein, B. Berkman, D. Brock, R. Cook-Deegan, S. M. Fullerton, H. Greely, et al., "Broad Consent for Research With Biological Samples: Workshop Conclusions," *American Journal of Bioethics* 15, no. 9 (2015): 34–42.

73. European General Data Protection Regulation, *available at* <https://www.eugdpr.org/> (last visited January 24, 2018).

74. Swiss Federal Department of Justice and Police "Den Datenschutz verbessern und den Wirtschaftsstandort stärken" *available at* <https://www.ejpd.admin.ch/ejpd/de/home/aktuell/news/2017/2017-09-150.html> (last visited January 24, 2018).

75. KPMG, "EU General Data Protection Regulation Ratified," (2016), *available at* <https://assets.kpmg.com/content/dam/kpmg/pdf/2016/05/EU-General-Data-Protection-Regulation-ratified-18-04-2016.pdf> (last visited January 24, 2018).

76. Laurie, *supra* note 52.

77. E. Vayena, U. Gasser, A. Wood, D. R. O'Brien, and M. Altman, "Elements of a New Ethical Framework for Big Data Research," *Washington & Lee Law Review* 72, no. 3 (2016): 420-441, *available at* <http://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/5/> (last visited January 24, 2018).