

By Alessandro Blasimme, Marta Fadda, Manuel Schneider, and Effy Vayena

DOI: 10.1377/hlthaff.2017.1558
HEALTH AFFAIRS 37,
NO. 5 (2018): 702-709
©2018 Project HOPE—
The People-to-People Health
Foundation, Inc.

Data Sharing For Precision Medicine: Policy Lessons And Future Directions

Alessandro Blasimme is a senior researcher in the Department of Health Sciences and Technology, Swiss Federal Institute of Technology—ETH Zürich, in Switzerland.

Marta Fadda is a postdoctoral researcher in the Department of Health Sciences and Technology, ETH Zurich.

Manuel Schneider is a PhD candidate in the Department of Health Sciences and Technology, ETH Zurich.

Effy Vayena (effy.vayena@hest.ethz.ch) is professor of bioethics in the Department of Health Sciences and Technology, ETH Zurich.

ABSTRACT Data sharing is a precondition of precision medicine. Numerous organizations have produced abundant guidance on data sharing. Despite such efforts, data are not being shared to a degree that can trigger the expected data-driven revolution in precision medicine. We set out to explore why. Here we report the results of a comprehensive analysis of data-sharing guidelines issued over the past two decades by multiple organizations. We found that the guidelines overlap on a restricted set of policy themes. However, we observed substantial fragmentation in the policy landscape across specific organizations and data types. This may have contributed to the current stalemate in data sharing. To move toward a more efficient data-sharing ecosystem for precision medicine, policy makers should explore innovative ways to cope with central policy themes such as privacy, consent, and data quality; focus guidance on interoperability, attribution, and public engagement; and promote data-sharing policies that can be adapted to multiple data types.

Precision medicine promises to create new therapies based on individuals' characteristics and lifestyles, improve health monitoring, reduce the use of ineffective drugs, and devise better public health interventions.^{1,2} No health care system can afford to ignore such possibilities. This explains why tailoring interventions to the real needs of patients—the ultimate goal of precision medicine—is receiving ample public support in countries like the US, the UK, and Switzerland.³⁻⁶ Notably, progress in precision medicine depends on the possibility of collecting, linking, and reusing large amounts of molecular, clinical, phenotypic, and lifestyle data on as many people as possible.⁷⁻¹¹ This implies that data sets should be accessible to multiple research groups and that data should be shareable.

Data sharing encompasses both transferring copies and enabling reuse of data. Access to

shared data can be open (if data are publicly available to all) or controlled (if permission must be sought to use a given data set). Sharing data refers to sharing both primary data, as in the case of the human genome sequences that were released right after assemblage, and secondary data that have already been used by those who originally collected and curated them.

Various models of data sharing exist. Yet how best to promote and govern the responsible sharing of human data for research and clinical purposes has long been a major policy challenge aggravated by the growth of private-sector data collection activities. Over the past two decades many stakeholders have produced a plethora of guidelines for data sharing that aim to address thorny issues around privacy, data security, and interoperability.¹² The resulting policy landscape consists of an extensive network of policy initiatives mobilizing a variety of principles and concerns and reaching out (with varying potential to

produce a tangible effect) to a heterogeneous array of publics.

Despite the proliferation of data-sharing guidelines, laments about data being insufficiently accessible are ever more frequent.¹³⁻¹⁶ This inaccessibility means that the often-invoked “cultural shift” needed to align data sharing with the needs of medical science has not occurred.^{12,17} Why this shift has not happened is a question that needs to be investigated, as its answer can inform the development of more effective policies. We began to explore this question by analyzing the data-sharing guidelines that have dictated practice. In what follows, we report the results of a comprehensive analysis of such guidelines issued over the past two decades by different types of organizations. To our knowledge, this is the broadest review of data-sharing policies conducted to date.

Our findings show a substantial fragmentation of the policy landscape around research data sharing. In light of that, we suggest that to favor a more efficient data-sharing ecosystem for precision medicine, policy makers should explore innovative ways to cope with central themes such as privacy, consent, and data quality; focus guidance on interoperability, attribution, and public engagement; and implement guidance that can be adapted to multiple data types.

Study Data And Methods

We analyzed 230 policy documents from the period 1996–2017 that provided guidance on data sharing (for an overview of the databases used for the search, a list of the included organizations, and our search strategy, see online appendixes 1, 2, and 3, respectively).¹⁸ We reviewed guidelines and best practices, official statements, policy analyses, ethics frameworks, and expert reports and declarations. The search, performed in September 2017, retrieved 205 eligible documents, to which we added 25 that we were previously aware of but that our search strategy did not uncover (see appendix 4).¹⁸

ANALYSES We used two methodological approaches. First, we employed a network analysis approach to explore the structural features of the data-sharing policy landscape and to map relationships among the included documents.^{19,20} Network analysis is a method used to generate explanatory hypotheses regarding social phenomena (in our case, the insufficiency of data sharing) by reconstructing the relations between participants in a given activity.²¹

We thus experimentally obtained and represented the network formed by the data-sharing guidelines as discrete objects on the World Wide Web.²² We studied the shape (path structure) of

this network to determine whether insufficient data sharing could be traced back to structural features of the policy network formed by the guidelines. To this end, we assessed the connectedness of the network against a baseline reference network (see appendix 5).¹⁸ The rationale for the use of this method is the assumption that network interconnectedness (or the lack thereof) is an indirect indication of a concerted policy effort around a given objective (in our case, promoting data sharing). We therefore used network analysis as a heuristic tool to investigate a recalcitrant policy issue and generate hypotheses to explain the source of policy ineffectiveness.

To corroborate the indications obtained via the network analysis, we also conducted a qualitative content analysis of the policy documents. We initially analyzed a subsample of fifty documents and inductively generated twenty-one meaningful categories or themes that could be applied for a later analysis of the entire sample (see appendix 6).¹⁸ Our categories identified normative claims (for example, “x should do a” and “y should be promoted”) that we refer to as policy themes (see appendix 7).¹⁸

We coded the full text of each document with the aid of qualitative research software (NVivo, version 11 pro) using our policy themes. Agreement among two of the coders was at least 95 percent in 98.2 percent of the cases (and kappa was at least 0.7 in 75 percent of the cases), which indicates the robustness of the inductively generated policy themes. We then developed an index to quantify the frequency of each policy theme in our sample. The index accounts for the number of times a reference to each policy theme would appear in relation to the total number of coded words in the documents. The result is then multiplied by 10,000 for better readability. For example, an index value of 1.83 for the theme of privacy indicates that privacy has been referred to 481 times. This index allowed us to look at how often each policy theme appeared within each type of document (clustered by the type of data each document was about) and within the documents of each type of organization.

LIMITATIONS One limitation of our study was the absence of an exhaustive repository of policy documents. In particular, bibliographic databases included only a small portion of the relevant documents. Nevertheless, our approach allowed us to analyze considerably more documents than similar studies have done.

A second limitation stemmed from the volatility of web-based content. For instance, the URLs leading to the included files might change over time. So as not to compromise reproducibility, we retained only PDF documents, which we will share with other researchers upon request.

Study Results

The 230 documents we analyzed covered a wide spectrum of data types, including health-related (30.9 percent), general research (20.4 percent), human genetic or genomic (20.4 percent), biobank (9.6 percent), scientific publication (2.2 percent), public health (2.2 percent), clinical trial (1.7 percent), and proteomic (0.4 percent) data. The largest share of the documents came from scientific societies (31.3 percent), followed by national public policy organizations (16.5 percent) and research funders (16.1 percent). Most documents were issued by organizations based in the UK (29.6 percent) or the US (18.2 percent) or by international organizations (18.2 percent) (see appendix 8 for an overview of the sample's characteristics).¹⁸

The data-sharing policy network²³ is considerably less dense than the reference network (see appendixes 9–11).¹⁸ In a perfectly connected network—one in which every node is connected to all other nodes—none of the nodes is more connected than the others, and the number of strongly connected components is equal to 1

(that is, the whole network is the most strongly connected component of itself). Our analysis revealed that the network under study had twenty-two times as many highly connected components compared to the reference network (appendixes 9 and 12).¹⁸ Considered together, these findings show that the policy network is rather loosely connected, which suggests a possible fragmentation of the policy landscape.

The content analysis of the documents confirmed this initial observation. For the past two decades the policy discourse around data sharing has been mobilizing a diverse variety of themes (exhibit 1 and appendix 6).¹⁸ However, the frequency with which the themes appeared in our sample revealed that not all of them have been equally emphasized.

Three themes were referred to much more frequently than others: data subjects' autonomy, data subjects' privacy, and data quality and curation (exhibit 1). Although these themes have been perceived as major concerns that merit special attention and guidance, not all organizations have given them equal consideration.

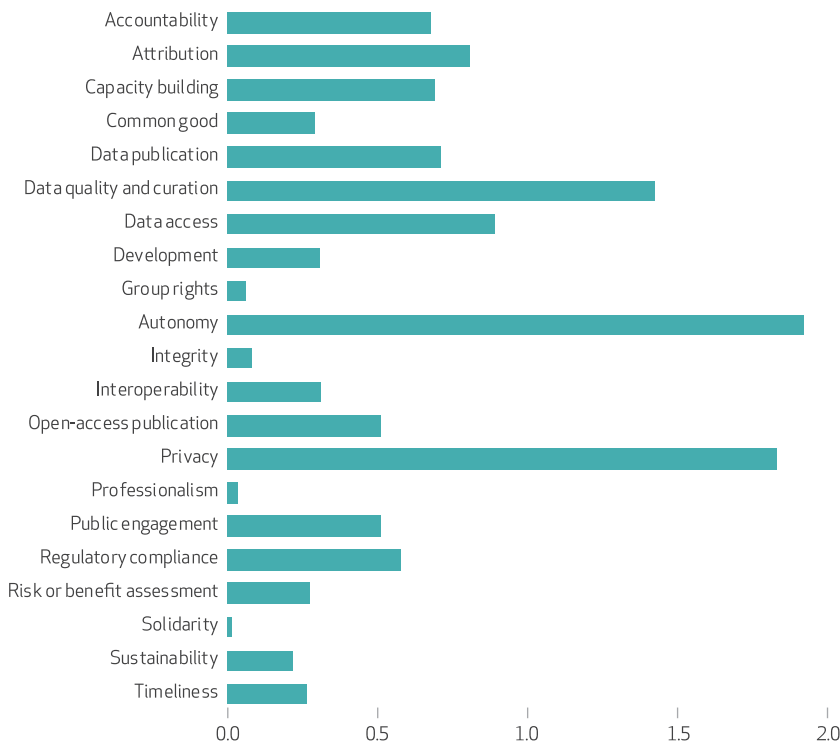
Almost all of the documents in our sample acknowledged the importance of data subjects' autonomy. This refers predominantly to procedures of informed consent as a precondition of collecting personal data and allowing their further uses. While other framings of autonomy are present (for example, data subjects' right to receive information generated from their data), informed consent is by far the most common. The type of organization that stressed this theme the most was research funders (exhibit 2). All other organization types emphasized autonomy as well, but not to the same degree as research funders (standard deviation: 0.33).

The policy theme of privacy appeared almost as frequently as that of autonomy in our sample (exhibit 1). This reveals a consensus about the need to balance the maximization of data availability with protecting data subjects' fundamental right to privacy. Research funders stressed privacy the most, followed closely by governments and governmental organizations (exhibit 2). Interestingly, however, this theme is the one that shows the highest variability in frequency across the different types of organizations. The standard deviation for privacy (0.76) was the highest that we encountered in our analysis.

The third most prominent policy theme, data quality and curation, captures recommendations addressing the need for sharable data to be reliable, complete, easily findable, and accompanied by sufficient metadata. Absent those characteristics, data sharing would have limited utility. Research funders emphasized this theme more than any other type of organization, but

EXHIBIT 1

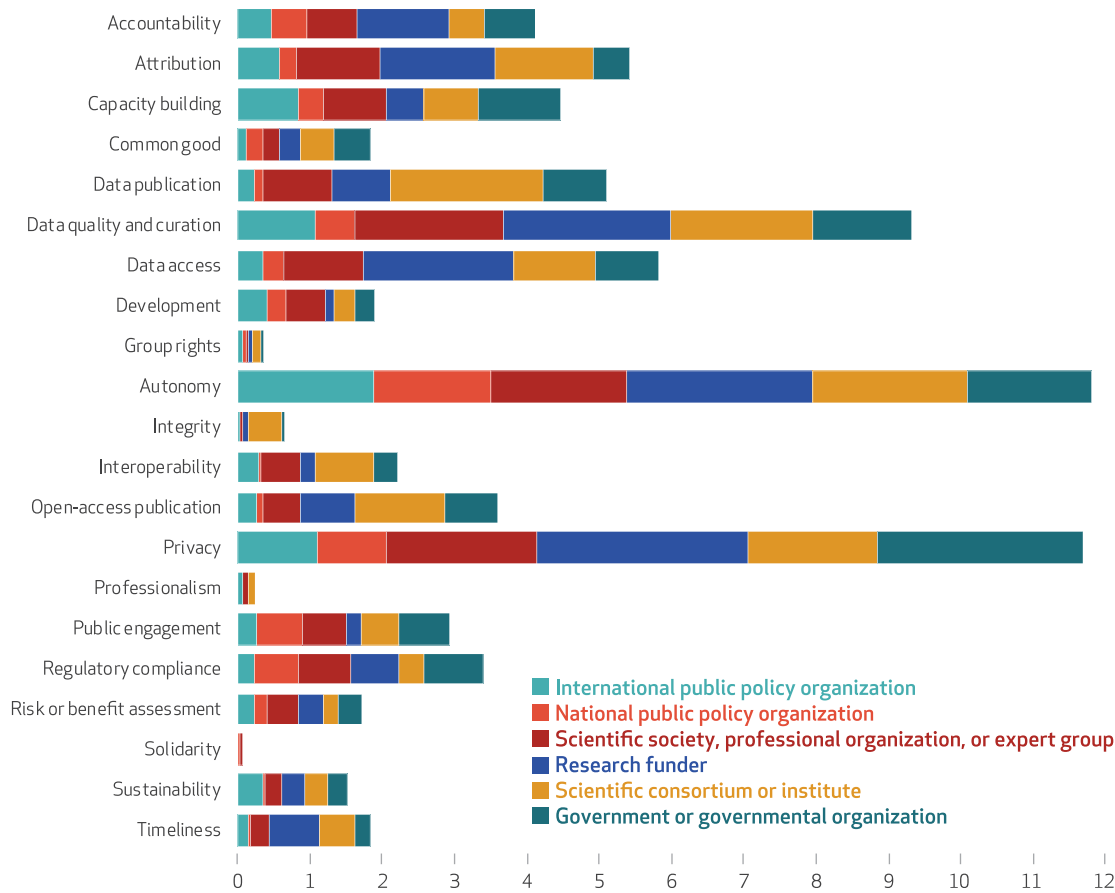
Frequency of references to the policy themes in the documents analyzed



SOURCE Authors' analysis using NVivo qualitative analysis software showing the policy themes' frequency across all documents. **NOTES** The sample consisted of 230 documents that provided normative guidance on data sharing. Frequency is measured by use of an index that accounts for the number of times a reference to each policy theme would appear in relation to the total number of coded words in the documents. The result is then multiplied by 10,000 to obtain an index ranging between 0 and 10.

EXHIBIT 2

Frequency of references to the policy themes, by organization type



SOURCE Authors' analysis using NVivo qualitative analysis software comparing the policy themes' frequency and the type of organization. **NOTES** The sample is explained in the notes to exhibit 1. Appendix 15.2 contains a list of all data used to generate this figure, and appendix 17 contains the mean and standard deviation for each theme (see note 18 in text).

the two types of scientific organizations also paid substantial attention to it (exhibit 2). However, despite the prominence of the theme, different types of organizations have covered it rather unevenly (SD: 0.61).

Other policy themes received less but still a fair amount of attention: data access (maximizing data reuse), attribution (finding ways to duly recognize and reward the work of data collectors and curators who make data available), data publication (publishing research data along with scientific papers), open-access publication of scientific works, capacity building (increasing knowledge regarding data science and data management), and accountability (creating robust oversight mechanisms) (exhibit 1). Interestingly, regulatory compliance (complying with existing regulations) and public engagement (fostering such engagement) fall close to the average frequency value (0.59) and received quite similar amounts of attention across organization

types (SDs: 0.21 and 0.2, respectively) (exhibit 2). Finally, the following handful of policy themes received only very limited attention: solidarity (appeals to that), professionalism (adherence to professional codes of conduct), group rights (considering the interests of local communities and populations in data-sharing practices), and integrity (framing data sharing as a matter of scientific integrity).

Furthermore, we observed that the most frequently addressed themes were those that were more unevenly cited in documents covering different types of data (see appendix 13).¹⁸ For example, data quality and privacy have the second- and third-highest standard deviations in the sample (2.98 and 2.37, respectively), with autonomy being only slightly more homogeneously distributed than these two (SD: 1.49) (see appendix 13).¹⁸

After checking for multicollinearity, we applied simple linear regression analysis to test

for whether the type of data discussed and the type of organization issuing the documents significantly predicted the frequency of the three most covered themes. Neither data type nor organization type was found to be significant predictors of data quality and curation (R^2 : 0.008; $F(2,227)$: 0.896; $p = 0.410$), privacy (R^2 : 0.001; $F(2,227)$: 0.075; $p = 0.928$), or autonomy (R^2 : 0.016; $F(2,227)$: 1.806; $p = 0.167$) (see appendix 14).¹⁸ Furthermore, not even the countries of the organizations were correlated significantly with data quality and curation ($F(13,216)$: 0.713; $p = 0.750$), autonomy ($F(13,216)$: 1.338; $p = 0.193$), or privacy ($F(13,216)$: 0.713; $p = 0.138$) (see appendix 14).¹⁸

Our analysis showed considerable fragmentation of the policy landscape: The network formed by the policy documents is poorly interconnected, not all policy themes are equally represented, different categories of organizations give different emphasis to the same themes, and documents addressing a given data type do not give similar emphases to the various themes.

Policy Implications

Although there might be multiple reasons why data sharing has not been achieved to the expected degree, it is highly probable that guidelines and policy approaches played an important role in what has or has not happened. In particular, the guidance we studied, despite its abundance, has not resulted in a cohesive system of incentives able to reconcile the interests and expectations of different stakeholders. The fragmentation that we observed, the emphasis on some specific themes, and the relative neglect of others led us to propose orienting future guidance in three directions described below.

DEVELOPING INNOVATIVE POLICY TOOLS The guidelines we analyzed focused primarily on the themes of autonomy (mainly through informed consent), privacy, and data quality and curation. While there is consensus that these themes are central to responsible data sharing, they also have been identified as typical barriers standing in the way of data sharing.¹⁵ One possible way to overcome such barriers is to explore innovative ways of addressing them.

► **AUTONOMY:** For decades, the protection of the autonomy of research participants and data donors has relied on informed consent. Originally designed as a form of autonomous authorization for research activities or medical acts,²⁴ informed consent has progressively acquired numerous other functions, including that of specifying individual preferences for the reuse of data and the return of incidental findings.²⁵ This has led to a functional overload.²⁶ While

The guidance we studied has not resulted in a cohesive system of incentives able to reconcile the interests and expectations of different stakeholders.

a necessary condition for secondary data use, informed consent as currently practiced does not offer the granularity that is required for data donors to exert meaningful control—particularly in the case of the multifarious data types required for precision medicine. New digital consent technologies relieve data donors from making complex up-front decisions; this enables case-by-case deliberations throughout the various uses of the data.^{27,28} Such tools could be promoted through pilot initiatives and ad hoc regulatory guidance in the context of national precision medicine initiatives.

► **PRIVACY:** In a similar vein, technological solutions could also be adopted to boost data privacy. For example, advanced cryptography and distributed ledger technologies such as the blockchain^{29,30} recently came to the fore as promising methods of protecting health data more effectively.^{31–35} Implementing such privacy-preserving technologies requires targeted public investment and the development of technical requirements for data exchange platforms to ensure regulatory compliance. The foreseeable payoff of such efforts is the growth of data infrastructures for precision medicine that can streamline data sharing, automate and increase data traceability, and reduce the risk of data breaches and misuse.

► **DATA QUALITY AND CURATION:** The role of data creators and curators is key for ensuring the quality and therefore the usability of data sets. Data-tagging systems have long been proposed as a way to trace credit for data curation “where credit is overdue.”³⁶ Yet such systems have struggled to establish themselves, as is demonstrated by the infrequent use of data-set indices or identification numbers.³⁷ In light of this, different

Our analysis shows that data-sharing guidance has so far not used levers across the full spectrum of relevant policy themes and impediments.

types of incentives need to be explored. One approach could be the creation of data infrastructures with a reciprocity-oriented design whereby, in order to get access to a database, researchers accept to contribute to its maintenance, curation, and development. Researchers could fulfill such a requirement either by contributing new data or by performing quality-related tasks for the database. A model of this kind can be incentivized in one of two ways: by insisting on the value of reciprocity in data-sharing guidance, and by introducing this feature up front in the design of new data infrastructures created for national flagship research initiatives in precision medicine.⁶

RECALIBRATING POLICY EMPHASIS Our analysis shows that data-sharing guidance has so far not used levers across the full spectrum of relevant policy themes and impediments. In particular, some important themes—interoperability, attribution, and accountability—have received comparatively limited consideration. Yet these policy themes have been identified as key enablers of data sharing^{12,37} and public trust in data-driven research.³⁸

Future guidance should therefore redistribute the policy emphasis across a broader set of policy themes to compensate for the observed fragmentation.

► **INTEROPERABILITY:** Data infrastructures for biomedical research should be interoperable—that is, ready to integrate multiple data formats (including data aggregated and uploaded by data subjects themselves) and able to seamlessly exchange data with each other. To date, however, efforts to increase interoperability have had only limited success. Lessons learned in the clinical field could prove valuable in the policy context. For example, linking reimbursement of accountable care organizations to quality and outcomes has been a potent incentive to

improve the interoperability of electronic health records.³⁹ Similarly, the adoption of interoperability standards is likely to increase data access requests. To incentivize the implementation of interoperable data standards, research funders could reward scientific institutions on the basis of the number of data access requests they grant to other research organizations.

► **ATTRIBUTION:** As to issues of fair attribution, the above-mentioned reciprocity-based approach to data reuse, coupled with the reward system we have just described, would enable data curators to be traceable. Yet to incentivize data citation practices, novel career paths for data curators would have to be introduced in research institutions. Such paths should include recognition of new skills and reward the role curators play in data sharing.

► **ACCOUNTABILITY:** Accountability has received increased attention in the discourse about responsible data processing.^{40,41} Guidance on how to streamline processes such as conducting impact assessments for privacy⁴² or data protection⁴³ could greatly improve accountability. Participatory data governance mechanisms also hold promise. Data subjects' involvement in decisions regarding data management would fulfill one of the aims of accountability—that is, to foster transparency and public trust in the use of data. This is particularly relevant for subjects who are enrolled in precision medicine studies based on group membership (for example, they belong to genetically isolated populations, disease groups, ethnic groups, minority groups, or specific communities). Offering subjects meaningful levels of control over data access and distribution practices includes ensuring that they are represented on data access committees and oversight bodies^{44,45} and can go as far as empowering subjects or communities to aggregate their own data from disparate sources and make them available for research under conditions they directly control.²⁷

Interestingly, participatory mechanisms are gaining traction in precision medicine as a way of more meaningfully considering the relevant interests of data subjects.^{3,46,47} Policy makers should therefore capitalize on the growing interest in public participation in precision medicine and set up participant-centered data governance models.

PROMOTING COMPREHENSIVE GUIDANCE Our analysis has indicated that the policy landscape around data sharing is fragmented. We have also shown that existing guidance tends to set apart different data types for regulatory purposes. Yet different guidelines treat the same data types in dissimilar ways. The resulting system of guidance has been undermined by this fragmented

approach, which may have contributed to a sub-optimal level of data sharing. Appeals for international policy harmonization have long been put forward, but this harmonization has fallen short of expectations.⁴⁸ This complicates data-centered research activities such as precision medicine, which require access to diverse data types simultaneously. For some more novel data types (such as unstructured data contributed by patients and data collected through mobile apps, wearable devices, or online social networks), there is limited, if any, guidance. Data-sharing guidance that applies across the board to multiple data types—thus focusing on the intended use of the data rather than their source—would facilitate access to heterogeneous data sources, from conventional to novel.⁴⁹ Likewise, oversight activities such as reviews of requests for data access should be able to accommodate such diversity. Future guidance to support precision medicine activities should therefore focus on creating the conditions necessary for the emergence of a coherent regulatory environment around the collection, use, and distribution of multidimensional health data.

Conclusion

We analyzed structural and content-related features of 230 documents from a wide range of international and national organizations from the period 1996–2017 that contained data-sharing guidance. We found that the documents concentrated on a restricted set of policy themes

and that there was substantial fragmentation of the policy landscape across specific policy organizations and contexts of use. Data-sharing guidance focuses on regulating areas in which stakeholders may have diverging interests—specifically, autonomy (typically through informed consent), privacy, and data quality and curation. With respect to these themes, different organizations may operate under fundamentally dissimilar sets of assumptions regarding the best way to treat the data they produce, collect, and process. Therefore, despite their reasonableness, existing policy approaches fall short of providing sufficient incentives for data sharing.

Further efforts by organizations and policy makers are needed to develop guidance to improve data sharing. Future policy initiatives will need to take stock of the already abundant guidance on data sharing. Our recommendations, based on three objectives, are a step in this direction. First, we suggest leveraging emerging technologies to streamline robust informed consent procedures and privacy-preserving data processing, and we propose the introduction of reciprocity-based data-access models to promote data quality. Second, we highlight some key policy themes—interoperability, attribution, and participatory governance—that have been on the radar of policy makers for quite some time but that must be reevaluated and prioritized. Finally, we suggest that data-sharing guidance for precision medicine must allow implementation across the widest possible range of data types. ■

This work was partly funded by the Swiss Academy of Medical Sciences (SAMW/ASSM) in the context of the Swiss Personalized Health Network and by the Swiss National Science

Foundation (Grant No. 157556, “Personalised Medicine, Ethics and Society”). The authors thank Agata Ferretti and Livia Modica for their valuable assistance in the preparation of

the manuscript and Elli Gourna for her involvement in earlier phases of the work.

NOTES

- 1 Khoury MJ, Bowen MS, Clyne M, Dotson WD, Gwinn ML, Green RF, et al. From public health genomics to precision public health: a 20-year journey. *Genet Med*. 2017 Dec 14. [Epub ahead of print].
- 2 National Research Council. *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease*. Washington (DC): National Academies Press; 2011.
- 3 Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med*. 2015;372(9):793–5.
- 4 Marx V. The DNA of a nation. *Nature*. 2015;524(7566):503–5.
- 5 Bahcall O. Precision medicine. *Nature*. 2015;526(7573):335.
- 6 Meier-Abt P, Lawrence AK, Selter L, Vayena E, Schwede T. The Swiss approach to precision medicine. *Swiss Medical Weekly [serial on the Internet]*. 2018 Jan 2 [cited 2018 Mar 16]. Available from: <https://smw.ch/en/op-eds/post/the-swiss-approach-to-precision-medicine/>
- 7 Kohane IS. Health care policy. Ten things we have to do to achieve precision medicine. *Science*. 2015;349(6243):37–8.
- 8 Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. *Bull World Health Organ*. 2018;96(1):66–8.
- 9 Jain SH, Powers BW, Hawkins JB, Brownstein JS. The digital phenotype. *Nat Biotechnol*. 2015;33(5):462–3.
- 10 Vayena E, Haeusermann T, Adjekum A, Blasimme A. Digital health: meeting the ethical and policy challenges. *Swiss Med Wkly*. 2018;148:w14571.
- 11 Elenko E, Underwood L, Zohar D. Defining digital medicine. *Nat Biotechnol*. 2015;33(5):456–61.
- 12 Sansone S-A, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, et al. Toward interoperable bioscience data. *Nat Genet*. 2012;44(2):121–6.
- 13 Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. *PLoS One*. 2011;6(6):e21101.
- 14 Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS One*. 2013;8(7):e67332.

- 15 Van Panhuis WG, Paul P, Emerson C, Grefenstette J, Wilder R, Herbst AJ, et al. A systematic review of barriers to data sharing in public health. *BMC Public Health*. 2014;14(1):1144.
- 16 Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. *JAMA*. 2016;315(12):1237–9.
- 17 Sejnowski TJ, Churchland PS, Movshon JA. Putting big data to good use in neuroscience. *Nat Neurosci*. 2014;17(11):1440–1.
- 18 To access the appendix, click on the Details tab of the article online.
- 19 Rogers R. *Digital methods*. Cambridge (MA): MIT Press; 2013.
- 20 Ackland R. *Web social science: concepts, data and tools for social scientists in the digital age*. London: Sage Publications; 2013.
- 21 Scott J, Carrington PJ, eds. *The SAGE handbook of social network analysis*. London: SAGE Publications; 2011.
- 22 Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Adar E, Hurst M, Finin T, Glance N, Nicolov N, Tseng B, eds. *Proceedings of the Third International Conference on Weblogs and Social Media [Internet]*. Menlo Park (CA): AAAI Press; 2009 [cited 2018 Mar 19]. p. 361–2. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009>
- 23 Borgatti SP, Mehra A, Brass DJ, Labianca G. Network analysis in the social sciences. *Science*. 2009;323(5916):892–5.
- 24 Faden RR, Beauchamp TL. *A history and theory of informed consent*. New York (NY): Oxford University Press; 1986.
- 25 Blasimme A, Moret C, Hurst SA, Vayena E. Informed consent and the disclosure of clinical results to research participants. *Am J Bioeth*. 2017;17(7):58–60.
- 26 Koenig BA. Have we asked too much of consent? *Hastings Cent Rep*. 2014;44(4):33–4.
- 27 Vayena E, Blasimme A. Biomedical big data: new models of control over access, use and governance. *J Bioeth Inq*. 2017;14(4):501–13.
- 28 Sage Bionetworks. *Better science together* [home page on the Internet]. Seattle (WA): Sage Bionetworks; c 2018 [cited 2018 Mar 19]. Available from: <http://sagebionetworks.org/>
- 29 Azaria A, Ekblaw A, Vieira T, Lippman A. Medrec: using blockchain for medical data access and permission management. In: Awan I, Younas M, eds. *Proceedings: 2016 2nd International Conference on Open and Big Data*. Washington (DC): Institute of Electrical and Electronics Engineers; 2016. p. 25–30.
- 30 Benchoufi M, Ravaud P. Blockchain technology for improving clinical research quality. *Trials*. 2017;18(1):335.
- 31 Dubovitskaya A, Xu Z, Ryu S, Schumacher M, Wang F. Secure and trustable electronic medical records sharing using blockchain [Internet]. Ithaca (NY): Cornell University Library; 2017 Aug 2 [cited 2018 Mar 20]. (arXiv:1709.06528v1). Available from: <https://arxiv.org/pdf/1709.06528.pdf>
- 32 Froelicher D, Egger P, Sousa JS, Raisaro JL, Huang Z, Mouchet C, et al. UnLynx: a decentralized system for privacy-conscious data sharing. *Proceedings on Privacy Enhancing Technologies*. 2017;2017(4):232–50.
- 33 Nikolaenko V, Weinsberg U, Ioannidis S, Joye M, Boneh D, Taft N. Privacy-preserving ridge regression on hundreds of millions of records. In: *Security and Privacy (SP), 2013 IEEE Symposium on*. IEEE; 2013. p. 334–48.
- 34 Raisaro JL, Choi G, Pradervand S, Colsenet R, Jacquemont N, Rosat N, et al. Protecting privacy and security of genomic data in i2b2. *IEEE/ACM Transactions on Computational Biology and Bioinformatics [serial on the Internet]*. [Cited 2018 Mar 19]. Available from: <https://infoscience.epfl.ch/record/229262/files/main.pdf>
- 35 Ackerman Shrier A, Chang A, Diakun-thibault N, Forni L, Landa F, Mayo J, et al. Blockchain and health IT: algorithms, privacy, and data [Internet]. Washington (DC): Department of Health and Human Services, Office of the National Coordinator for Health Information Technology; 2016 Aug 8 [cited 2018 Mar 19]. (White Paper). Available from: https://www.healthit.gov/sites/default/files/1-78-blockchainandhealthalgorithmsprivacydata_whitepaper.pdf
- 36 Credit where credit is overdue. *Nat Biotechnol*. 2009;27(7):579.
- 37 Thorisson GA. Accreditation and attribution in data sharing. *Nat Biotechnol*. 2009;27(11):984–5.
- 38 Kaye J, Heeney C, Hawkins N, de Vries J, Hoodington P. Data sharing in genomics—re-shaping scientific practice. *Nat Rev Genet*. 2009;10(5):331–5.
- 39 Washington V, DeSalvo K, Mostashari F, Blumenthal D. The HITECH era and the path forward. *N Engl J Med*. 2017;377(10):904–6.
- 40 Hunton and Williams LLP. *Accountability: a compendium for stakeholders* [Internet]. Washington (DC): Centre for Information Policy Leadership; 2011 Mar [cited 2018 Mar 19]. Available from: <http://informationaccountability.org/wp-content/uploads/Centre-Accountability-Compendium.pdf>
- 41 European Data Protection Supervisor. *Accountability on the ground: provisional guidance on documenting processing operations for EU institutions, bodies and agencies: summary* [Internet]. Brussels: EDPS; 2018 Feb [cited 2018 Mar 19]. Available from: https://edps.europa.eu/sites/edp/files/publication/18-02-06_accountability_on_the_ground_summary_en.pdf
- 42 Information Commissioner's Office. *Conducting privacy impact assessments: code of practice: Data Protection Act* [Internet]. Wilmslow (UK): ICO; 2014 Feb [cited 2018 Mar 19]. Available from: <https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf>
- 43 Katz J. The regulation of human research—reflections and proposals. *Clin Res*. 1973;21(4):785–91.
- 44 O'Doherty KC, Burgess MM, Edwards K, Gallagher RP, Hawkins AK, Kaye J, et al. From consent to institutions: designing adaptive governance for genomic biobanks. *Soc Sci Med*. 2011;73(3):367–74.
- 45 Winickoff DE, Jamal L, Anderson NR. New modes of engagement for big data research. *Journal of Responsible Innovation*. 2016;3(2):169–77.
- 46 Blasimme A, Vayena E. Becoming partners, retaining autonomy: ethical considerations on the development of precision medicine. *BMC Med Ethics*. 2016;17(1):67.
- 47 Blasimme A, Vayena E. “Tailored-to-you”: public engagement and the political legitimization of precision medicine. *Perspect Biol Med*. 2016;59(2):172–88.
- 48 Arzberger P, Schroeder P, Beaulieu A, Bowker G, Casey K, Laaksonen L, et al. Science and government. An international framework to promote access to data. *Science*. 2004;303(5665):1777–8.
- 49 Vayena E, Blasimme A. Health research with big data: time for systemic oversight. *J Law Med Ethics*. 2018;41(6):119–29.