2019

Biobanks as Innovation Infrastructure for Translational Medicine

W. Nicholson Price II
University of Michigan Law School, wnp@umich.edu

Publication Information & Recommended Citation

Price, W. Nicholson, II. "Biobanks as Innovation Infrastructure for Translational Medicine." In *Global Genes, Local Concerns: Legal, Ethical, and Scientific Challenges in International Biobanking*, edited by Timo Minssen, Janne R. Herrmann, and Jens Schovsbo, 42-55. Cheltenham, U.K.: Edward Elgar Publishing, 2019.

3. Biobanks as innovation infrastructure for translational medicine

W. Nicholson Price II[1]

Biobanks represent an opportunity for the use of big data to drive translational medicine. Precision medicine demands data to shape treatments to individual patient characteristics; large datasets can also suggest new uses for old drugs or relationships between previously unlinked conditions. But these tasks can be stymied when data are siloed in different datasets, smaller biobanks, or completely proprietary private resources. This hampers not only analysis of the data themselves, but also efforts to translate data-based insights into actionable recommendations and to transfer the discovered technology into a commercialization pipeline. Cross-project technological innovation, development, and validation are all more difficult when data are divided between different biobanks and other data repositories.

One way to conceive of biobanks and the big medical datasets they create and embody uses the lens of infrastructure: how can biobanks and their data serve as infrastructure to support later innovation? Some efforts already fit into this model; for example, the United States' Precision Medicine Cohort—now renamed All of Us—aims to create a large, uniform dataset to be used for widespread future research. Other biobank-related data efforts, like Myriad's dataset on BRCA1/2 genetic variations, still function as entirely private resources. Treating medical big data as infrastructure has implications for how they should be governed, and suggests advantages to centralized control and relatively broad access. More broadly, viewing biobank-related data as infrastructure would place them at a distinctly earlier point in the commercialization pipeline, serving more to facilitate later steps in translational medicine rather than being viewed as potentially commercializable products themselves.

This chapter is divided into two parts. In the first, I briefly describe big data in medicine: the sources of medical data, the promises of medical big data,

[1] For helpful comments and conversations, I wish to thank Ana Bracic, Rebecca Eisenberg, Brett Frischmann, and Timo Minssen. Rebecca Kaplan provided excellent research assistance. All errors are my own.

and a key challenge: data fragmentation. In the second, I discuss the role of biobanks in medical big data, focusing on their role in infrastructure for innovation and their potential for facilitating translational research.

1 BIG DATA IN MEDICINE

Big data has long been heralded as the next big revolution in health care—but that revolution has been relatively slow to arrive. Although data are constantly and increasingly generated from many sources of medical information, including research and samples associated with biobanks, those data are often fragmented into segments that are less useful than might be the whole. This section briefly describes the sources of medical data, the potential benefits of such data, and the challenge of fragmentation.

1.1 Sources of Medical Data

Big health data come in many forms. The most traditional, of course, are the health records generated in routine medical encounters, and now captured in electronic health records (EHRs).[2] These include doctors' notes, test results, patient medical history, diagnoses, and other medical information.[3] Insurance claims records, raw diagnostic testing data, and prescription records increase the picture of medical data. Less traditional, but increasingly a part of the picture, are the health-related data collected by wearable devices (medical or otherwise), including fitness trackers, insulin monitors, and smartphones.[4] Finally—and especially important in the context of this work—research data and patient samples, while only available for a subset of patients, provide extraordinarily deep data for that set. They often aim to provide an especially complete set of medical information for a particular patient because of the potential to answer questions that might arise later.

1.2 Promised Benefits

Big data promise substantial benefits for the health system. In the short term, they are supposed to help drive efficiency in health systems, and to show patterns of care, how practices can be improved, and the like.[5] But the bigger

---

[2] *See generally* Sharona Hoffman, Electronic Health Records and Medical Big Data: Law and Policy (1 edition ed. 2016).

[3] *Ibid.*

[4] *See* W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 Mich. L. Rev. 421 (2017).

[5] *Ibid.*

promise comes from future potential for innovation. Precision medicine promises to tailor care to individuals based on their individual characteristics. Some such relationships can be painstakingly and explicitly derived, leading to hypotheses testable through classical clinical trials. Other, more challenging methods rely on using truly vast sets of data and turning machine-learning algorithms loose on those datasets to find complex, implicit patterns.[6]

1.3 The Challenge of Data Fragmentation

Data fragmentation is a tremendous barrier to realizing the potential for medical big data, and the barrier on which this article focuses.[7] The promise of big data depends on linking data from multiple sources for an individual patient, and on linking data across many patients to determine useful patterns to direct innovation and care. Ideally, the available datasets would include comprehensive information for a broad set of patients. Unfortunately, data are generated by different sources and are often difficult to reunite. Primary care physicians, specialists, and others involved directly in care may maintain their own records, which are only sometimes linked. And even when data are linked across the spectrum of care, they are often unconnected from those data generated outside the context of care. Other data arise from research contexts, and may or may not be linked to clinical care data.[8] Biobanks may acquire both sources of data, as they can acquire both patient health records (or some fraction of such records) and data from research studies. Finally, some data arise from sources far from the health system, such as wearable devices or internet searches; these are currently unlikely to be linked with other health records save through the action of the patient in question. This fragmentation is exacerbated over time, as patients switch doctors, insurers, pharmacies, and wearable technologies, and join or drop out of research studies. Even a patient's primary

---

[6] *See* W. Nicholson Price II, *Black-Box Medicine*, 28 Harv. J.L. & Tech. 419 (2015).

[7] Other barriers certainly exist, and I do not mean to downplay them here. Data quality is a substantial hurdle. *See*, *e.g.*, Sharona Hoffman & Andy Podgurski, *Big Bad Data: Law, Public Health, and Biomedical Databases*, 41 J. Law. Med. Ethics 56 (2013). Other technological hurdles include storage and analyses of data. *See*, *e.g.*, Niels Peek et al, *Technical Challenges for Big Data in Biomedicine and Health: Data Sources, Infrastructure, and Analytics*, 9 Y. B. Med. Inform. 42 (2014). For a description of other barriers in large scale observational research, *see*, *e.g.*, Rebecca S. Eisenberg & W. Nicholson Price II, *Promoting Healthcare Innovation on the Demand Side*, 4 J. L. & Bioscience 3, 23–39 (2017).

[8] Sometimes legal barriers limit integration of research data into clinical care records, as when in the US a laboratory performs research but is not approved under the Clinical Laboratory Improvement Amendments (CLIA) to perform clinical testing.

care records can become time-fragmented if the patient is not diligent about having records transferred from one doctor to the next—and even if the patient is diligent, the lack of compatibility between different electronic health records may frustrate the merging of information.[9]

In addition to fragmentation of data within the records of an individual patient, of course, there is tremendous segregation of data from different patients. Doctors, hospitals, insurers, and others have little individual incentive to make their data available to those who would combine them into larger datasets—and in fact may be prohibited from doing so by privacy and security rules in many contexts.[10]

This fragmentation of data hinders the goals of big data in medicine, and limits the insights that can be derived.[11] Less comprehensive datasets limit the relationships that can be identified, and may lead to biased outcomes.[12] While overcoming fragmentation is not the only challenge to the use of big data to drive both basic and translational medicine, it is a substantial hurdle.

2 BIOBANKS, INNOVATION INFRASTRUCTURE, AND TRANSLATIONAL MEDICINE

Biobanks create possible avenues to use big data better in the context of translational medicine. Biobanks are well positioned to gather, generate, and store medical big data. By bridging the gap between basic research and real-world patient phenotypic samples and data, they can facilitate the translation of laboratory insights into clinical practice. In so doing, they play an infrastructure role, both *for* and *of* big data in health. By an infrastructure *for* data, I mean that biobanks can provide resources to store, transfer, analyze, and otherwise use data. But biobanks can also help create an infrastructure *of* data—that is, the data that biobanks create and store are themselves infrastructure for translational innovation. This section briefly addresses each of these issues.

---

[9] *See*, *e.g.*, Andy Kessler, *Siri, Am I About to Have a Heart Attack*, Wall St. J. (Jan. 9, 2017), www.wsj.com/articles/siri-am-i-about-to-have-a-heart-attack-1484007412 (noting challenges of EHR interoperability for medical big data and noting that Epic Systems, the leading provider of EHR systems, "appear[s] to be the leading obfuscator when it comes to transferring records and interoperability").

[10] *See*, *e.g.*, Eisenberg & Price, *supra* note 6, at 34–9 (discussing the challenges to data integration posed by the Health Information Portability and Accountability Act (HIPAA)).

[11] *See* W. Nicholson Price II, *Risk and Resilience in Health Data Infrastructure*, 16 Colorado Sci. & Tech. L. Rev. 65 (2017).

[12] *See* Price, *Black-Box Medicine*, *supra* note 5, at 430–2 (2015) (describing the desirability of large datasets to identify complex relationships).

2.1 Biobanks and Medical Big Data

Biobanks occupy a special role in the universe of medical big data for at least three reasons. First, at least some biobanks collect data from at least two spheres of data—health care data and research data—that are often held separately. Biobanks are in the business of collecting both samples and data from patients;[13] the first are substantially less useful without the second. Thus, biobanks collect patient medical information along with samples.[14] To the extent that biobanks acquire research results about those samples, they have the advantage of aggregating both medical and research data about an individual. Researchers who actually analyze the samples also acquire both types of information, of course, but only for the patients in their own studies, while biobanks can potentially join information about many more patients represented in their collections. Meanwhile, in the process of acquiring and processing samples, biobanks may perform many analyses outside whatever research protocol was specified for gathering the sample in the first place; for instance, biobanks may sequence genomes, quantify mRNA populations, measure metabolite and/or protein levels, and histologically classify samples.

Second, because biobanks maintain collections of biological specimens, there exists the potential for performing currently unplanned analyses.[15] Uniquely among repositories of patient information, biobanks have the capacity to generate significant amounts of new data without acquiring it from individuals, by reanalyzing samples using new technology. To take an obvious example, consider a collection of tumor samples gathered throughout the course of several decades. For most of that time, the samples would not have been genetically analyzed because the technology was not available. But now, the entire set of samples could be genetically sequenced and the resulting sequence data could be linked to tumor pathology and other medical information about the patients that the biobank recorded. Sometimes this approach can create controversy, as with genetic analyses of blood spots collected from newborn infants. Ideally, such analyses are facilitated both by broad upfront

---

[13] Indeed, some have suggested that merely by standardizing the collection of data and patients within the catchment of a biobank, patient care may already be improved. Conor M.W. Douglas & Philip Scheltens, *Rethinking Biobanking and Translational Medicine in the Netherlands: How the Research Process Stands to Matter for Patient Care*, 23 Eur. J. Hum. Genet. 736 (2015).

[14] *See, e.g.*, Timo Minssen & Jens Schovsbo, *Legal Aspects of Biobanking as Key Issues for Personalized Medicine and Translational Exploitation*, 11 Pers. Med. 497 (2014).

[15] *See, e.g.*, Gerardo Botti et al, *Tumor Biobanks in Translational Medicine*, 10 J. Transl. Med. 204, 204 (2012).

consent (or other models that permit ongoing consent) and by the addition of more recent health data from the individual, where available.

Third, biobanks—or at least some fraction of them—already have as part of their mission a role as the repository for information, whether embedded in biological specimens or found in biological data. They are created with the idea that they will collect samples and make those samples available to future researchers, along with associated data. Thus, biobanks already provide infrastructure for biomedical innovation. This role is acknowledged explicitly in some cases. For instance, the Austrian-headquartered Biobanking and BioMolecular resources Research Infrastructure (BBMRI)[16] is a "distributed research infrastructure of biobanks and biomolecular resources," which aims to connect researchers and biobanks and "facilitate the use of samples/data collected in Europe for the benefit of human health."[17] This raises the question: how exactly can biobanks help enable the use of samples and data for human health? A longtime answer is that biobanks can provide resources that are useful for basic research. But biobanks can also facilitate research later in the pathway.

2.2 Biobanks and Translational Medicine

Biobanks are a key resource for developing translational medicine. They help make the jump from basic research discoveries to the phenotypic reality of patient populations represented by samples and data.[18] In a meaningful sense, this is because biobanks themselves straddle the divide between basic and clinical research; they are established as tools to help initial research, but do so by collecting large amounts of real-world samples and data.

Biobanks can facilitate translational medicine in several ways. For instance, a basic lab discovery might identify a gene with potentially significant clinical impact because it encodes a protein that might be a potential drug target. Biobank samples and data can thus help demonstrate whether relevant gene variants are present in patients in the represented population, and can demonstrate real-world correlations with the disease of interest.[19] Biobanks can similarly be used to identify biomarkers to be used in drug development, and later to validate and quantify those same biomarkers.[20] Such biomarkers can

---

16 BBMRI, *Frequently Asked Questions*, www.bbmri-eric.eu/faq/.

17 *Ibid.*

18 Minssen & Schovsbo, *supra* note 13.

19 *See* Botti et al, *supra* note 14.

20 *See* Arndt A. Schmitz, *Potential of Biobanking in Translational Medicine*, Presentation, HandsOn: Biobanks (Helsinki, Finland, 2014), *available at* http://

be prognostic (predicting the natural course of an illness),[21] predictive (helping identify a specific treatment), or pharmacodynamic (suggesting an optimal dose).[22] Finally, biobank data from patients who have already participated in clinical trials may be able to help stratify patients in those trials retroactively, and to develop new information about the already tested therapeutic agents.[23]

2.3 Biobanks as Data Infrastructure

Biobanks can serve an important role in providing data infrastructure for translational innovation in medicine. When I say data infrastructure, I mean both infrastructure *for* data—that is, resources for storing, collecting, and using data—and infrastructure *of* data—that is, the data themselves as infrastructure for later innovation.

Before getting into these two types of data infrastructure, it is worth being more explicit about what I mean by infrastructure. I principally adopt Brett Frischmann's characterization of infrastructure: (1) resources that can be "consumed nonrivalrously for some appreciable range of demand," which demand (2) is "driven primarily by downstream productive activities that require the resource as an input," where (3) those downstream productive activities result in a "wide range of goods and services, which may include private goods, public goods, and social goods."[24] The fact that infrastructure is widely usable for a broad range of outputs, some of which are public goods, implies that infrastructure is likely to be underprovided by private sources,[25] and also suggests that infrastructure resources are best kept relatively general to allow many uses rather than being specialized for one particular use.[26]

Biobanks are a promising source of infrastructure *for* data. By that I mean that they are designed to be repositories of samples for use by researchers, whether those researchers are at the early stages of discovery or later, in the process of translating fundamental insights into useful treatments that can be implemented in the clinic. Biobanks can serve a similar role—though

---

handsonbiobanks.org/documents/114074/129625/Schmitz_Biobanking_in_translational_medicine_HOBB2014.pdf/0dd2c1e9-298d-4af2-bc63-a34895442d7e.

[21] *See, e.g.*, Tobias M. Gorges & Klaus Pantel, *Circulating Tumor Cells as Therapy-Related Biomarkers in Cancer Patients*, 62 Cancer Immunol. Immunother. 931 (2013).

[22] *See* Schmitz, *supra* note 19. Biobanks can also supply samples not only to identify new biomarkers, but also to develop assays to measure those biomarkers. *Ibid.*

[23] *Ibid.*

[24] Brett Frischmann, Infrastructure: The Social Value of Shared Resources 61–2 (2012).

[25] *Ibid* at 15.

[26] *Ibid* at 65.

even more explicit—for data related to such samples, or even more broadly. They provide the physical resources—freezers and collection equipment for samples, computers and networks for data—that let these valuable resources be collected, stored, accessed, and used. They can also provide intangible infrastructural resources, such as protocols for sample and data collection, patient procedures, or even norms about collection and use.[27] These resources may be tied directly to the biobank, or to the umbrella organization as a parallel to the sample-driven physical biobank.[28] One such example may be found in New Haven, where Yale University hosts the world's largest genomic biobank.[29] The biobank consists of specimens and data from more than 500,000 participants in the ongoing Million Veteran Program.[30] The biobank itself will store and maintain the data, providing an infrastructure for those data.

But biobanks are also important in the creation of an infrastructure *of* data—by which I mean that they generate, maintain, and make accessible information which is itself infrastructure that provides resources for future innovation.[31] This goal may be explicit; the Yale biobank, for instance, is best "viewed as a long-term infrastructure project" providing support for current and future researchers, according to its codirector.[32] The Precision Medicine Cohort (now All of Us), formed as part of President Obama's Precision Medicine Initiative, similarly aims to develop a very large dataset that can be used to support future innovation.[33]

How might such a broad infrastructure project work best? And what does the conception of infrastructure for biobanks gain us? Ideally, infrastructure

---

[27] *See*, *e.g.*, Nat'l Comm. on Vital & Health Stats, Information for Health: The Strategy for Building the National Health Information Infrastructure 11 (2001) (defining infrastructure for health data very broadly).

[28] Although I argue the mission and funding of data infrastructure and sample storage are similar, I recognize that some resources and forms of expertise differ between the two functions.

[29] John D. Curtis, *Million Veterans Program Now World's Largest Genomic Biobank*, Yale School of Medicine News (Aug. 8, 2016), https://medicine.yale.edu/news/article.aspx?id=13225.

[30] *Ibid.*

[31] *See* OECD, Data-Driven Innovation: Big Data for Growth & Well-Being, 177–206 (2015) (applying an infrastructure model for big data generally); W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 Cardozo L. Rev. 1401, 1439–44 (2016) (describing an infrastructure model for medical big data to support the development of complex medical algorithms to direct treatment).

[32] Curtis, *supra* note 28.

[33] Francis S. Collins & Harold Varmus, *A New Initiative on Precision Medicine*, 372 N. Engl. J. Med. 793 (2015) (describing the precision medicine initiative); Eisenberg & Price, *supra* note 6, at 44 (describing the Precision Medicine Cohort as government-provided innovation infrastructure).

for and of data in the biobank context should be connected, interoperable, and accessible.[34] These ideals arise because of the nature of infrastructure in enabling a broad range of different users and uses,[35] and are closely related to the FAIR Guiding Principles for data management laid out in 2016: findability, accessibility, interoperability, and reusability.[36] In playing an infrastructural role, not only should biobanks ensure that data are findable and reusable, but they should also proactively link data to make them usable in many contexts and future studies.

Connection means that individual players—in this case, individual biobanks—should be connected to each other, sharing resources and data.[37] This helps make the available datasets bigger and more comprehensive, which in turn enables the study of more complex relationships or rare conditions.[38] In addition, connection helps ensure that biobanks as a group facilitate broad and varied uses rather than focusing on uses specifically tailored to a particular use or user.

Interoperability is a key enabler of connection. That is to say, if biobanks store their data in different, mutually incompatible formats, connection becomes much more challenging.[39] Such interoperability challenges are already a major concern in the context of electronic health records.[40] To the extent that biobanks create their own data structures, interoperability concerns can swamp the possibility of meaningfully connected data. Policy efforts should therefore encourage the use of compatible data formats to better enable

---

[34] For a broader description of several principles for data as infrastructure, *see* OECD, *supra* note 30, at 188 ff.

[35] *See* Frischmann, *supra* note 23, at 61–2.

[36] Mark D. Wilkinson et al, *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, 3 Scientific Data 160018 (2016).

[37] *See, e.g.*, Botti et al, *supra* note 14 (noting cooperation between biobanks allowing the study of rare cancers); OECD, *supra* note 30 (describing the need for connection).

[38] Examples of more comprehensive datasets—whether through centralized or distributed architecture—are becoming more common. The FDA's safety surveillance Sentinel system, for instance, relies on a distributed architecture where data are kept by their creators but are available for centralized querying. *See* Susan Forrow et al, *The Organizational Structure and Governing Principles of the Food and Drug Administration's Mini-Sentinel Pilot Program*, 21 Pharmacoepidemiol. Drug Saf. 12 (2012); Eisenberg & Price, *supra* note 6, at 41–4 (describing the use of possible use of Sentinel or Sentinel-like systems to promote healthcare innovation by payers).

[39] *See* OECD, *supra* note 30, at 192–94.

[40] *See* Eisenberg & Price, *supra* note 6, at 25–6 (describing interoperability challenges).

a data infrastructure.[41] For instance, the Minimum Information About Biobank data Sharing model (MIABIS), developed by the BBMRI, aims to create a new "bio-object infrastructure" within the EU.[42]

Finally, to provide infrastructure for and of data, those data must be accessible to researchers. Finding the correct model of accessibility is not easy.[43] The costs of access may be substantial, and funding such access either by surcharges on researchers using the information or by other public/private mechanisms each have their own challenges.[44] Funding based on fees to users is the most straightforward possibility, but risks privileging larger market incumbents over new entrants, and undermines the infrastructure model of biobanks.[45] This is true because of the varied nature of output goods from infrastructure goods; users that create downstream public goods are unwilling to pay for access to the resource at a socially desirable level.[46] Similarly, those developing infrastructural goods are unlikely to invest at a socially optimal level because they cannot capture the full benefits.[47] Myriad Genetics provides a useful example of this dynamic in action: Myriad keeps its vast trove of health and genetic information on women who have used its BRCAnalysis service to test for mutations in the BRCA1 and BRCA2 genes, linked to breast

---

[41] *See*, *e.g.*, U.S. Office of the National Coordinator for Health Information Technology, Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap (Draft) 10–11 (2015), www.healthit.gov/sites/default/files/nationwide-interoperability-roadmap-draft-version-1.0.pdf.

[42] Loreana Norlin et al, *A Minimum Data Set for Sharing Biobank Samples, Information, and Data: MIABIS*, 10 Biopreservation Biobanking 343 (2012). The development of the Data Sharing model has been a complex process. *See* Sakari Tamminen, *Bio-Objectifying European Bodies: Standardisation of Biobanks in the Biobanking and Biomolecular Resources Research Infrastructure*, 11 Life Sci. Soc. Policy 13 (2015).

[43] *See* Kathleen Liddell & Johnathon Liddicoat, *Open Innovation with Large Bioresources: Goals, Challenges, & Proposals*, University of Cambridge Faculty of Law Research Paper No 6/2017 (2017), *available at* https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2888871.

[44] *See* OECD, *supra* note 30, at 191–92.

[45] *See ibid* at 15, 18–19 (discussing different models of covering costs); *cf.* Barbara J. Evans, *Sustainable Access to Data for Postmarketing Medical Product Safety Surveillance under the Amended HIPAA Privacy Rule*, 24 Health Matrix 11 (2014) (noting the challenges to funding for access to data under Health Insurance Portability and Accountability Act requirements forbidding adequately charging for access, and suggesting the availability of higher fees).

[46] Frischmann, *supra* note 23, at 68–9.

[47] *Ibid.*

cancer.[48] While the company invested enough to create a valuable resource of genetic data and sequences for its own use, its exclusionary business model meant that other users could not access those samples or data for socially beneficial purposes such as better understanding or confirmatory testing.[49]

Funding from other private or public resources, such as general tax revenues, is thus likely a better solution in terms of enabling broad access.[50] Infrastructure goods typically create substantial spillovers—indeed, that is much of their purpose—bolstering the case for public funding.[51] But public funding raises political economy concerns of procuring the funding in the first place and of—arguably—leaving money on the table once innovations have been developed.[52] This latter concern is the obverse of the spillover benefit: spillovers are, by definition, uncaptured benefits, and while those are a classic benefit of public spending, they can also create friction and concerns about properly managing the public risk. Some approaches might try to blend the public and private funding model to resolve this tension, perhaps requiring different access fees for different types of data users.[53] Setting the appropriate balance in such a blended approach may bring its own difficulties.

Funding is not the only challenge; balancing data accessibility against privacy is nontrivial.[54] As described above, broader access is important for an infrastructural good to enable various downstream uses. Nevertheless, individuals regard health data as sensitive, and thus privacy concerns arise, not only from the perspective of participant buy-in but also to satisfy legal and policy requirements. Various access models have been proposed to lower privacy risks, including models drawing from the literature on sharing data

---

[48] *See*, *e.g.*, Robert Cook-Deegan et al, *The Dangers of Diagnostic Monopolies*, 458 NATURE 405 (2009).

[49] *Ibid.*

[50] Frischmann, *supra* note 23, at 94.

[51] *Ibid* at 14.

[52] *See* Rainer Warth & Aurel Perren, *Construction of a Business Model to Assure Financial Sustainability of Biobanks*, 12 BIOPRESERVATION BIOBANKING 389 (2014); Liddell & Liddicoat, *supra* note 42, at 18–19; *see* OECD, *supra* note 30, at 191–2.

[53] Liddell & Liddicoat, *supra* note 42, at 12–13 (noting that the 100,000 Genomes Project in the UK applies differential licensing and IP terms to different entities seeking to use the Project's data).

[54] *See* Roger A. Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 22 MICH. TELECOMM. & TECH. L. REV. 1 (2016) (describing the tradeoff between privacy of patient data and the ability to verify the quality of complex medical algorithms developed using those data); INST. OF MED., SHARING CLINICAL TRIAL DATA: MAXIMIZING BENEFITS, MINIMIZING RISK (2015), *available at* http://nap.edu/18998 (extensively analyzing the sharing of clinical data and suggesting different models)

from clinical trials.[55] Technical mechanisms may also help to allow data access without reducing privacy, though these come with their own hurdles to adoption.[56] Despite the challenges, access is key; without access, the data risk staying in isolated silos that help promote neither early stage innovation nor later translational work.

In the present, these goals of connection, interoperability, and access are far from a reality, at least in many instances; biobanks and the data they house are highly fragmented, limiting their value as infrastructure resources.[57] There are at least several hundred biobanks across the globe, and perhaps thousands, depending on methodology and definitions.[58] Different biobanks are and were founded for different purposes, and to serve different patients—who may have had different views about acceptable purposes for their samples' retention, and who may have consented to different types of future research. Nonprofit and for-profit biobanks may have different motives, but still often keep resources fragmented. For-profit biobanks are driven by competitive forces to keep their samples and data tightly siloed and unavailable to others. The contours of intellectual property rights, trade secrecy, and different regulatory rules for different types of data may unfortunately encourage data silos of that

---

[55] *See*, *e.g.*, Ford & Price, *supra* note 53, at 29–43.

[56] *Ibid.*

[57] Other problems arise from legal questions, in particular the issue of intellectual property rights and the desire for proprietary data. *See, e.g.,* Minssen & Schovsbo, *supra* note 13; Michiel Verlinden, Timo Minssen, & Isabelle Huys, *IPRs in Biobanking: Risks and Opportunities for Translational Research*, 2 Int. Prop. Quarterly 106 (2015). Those concerns will not be addressed here, other than to note that the desire to keep data and samples proprietary—or to exert strong intellectual property rights to limit future use or demand substantial compensation—cuts against the idea of biobanks as providing broadly accessible infrastructure for future innovation. *See* Liddell & Liddicoat, *supra* note 42. Toll roads may be useful to maintain a reasonable infrastructure, but excessive tolls slow the flow of useful traffic.

[58] *See* Gregory J. Boyer & Warren Whipple, *Biobanks in the United States: How to Identify an Undefined and Rapidly Evolving Population*, 10 Biopreserv Biobank 511 (December 2012) (finding hundreds of biobanks and acknowledging the uncertainty of the count); R. Jean Cadigan, Dragana Lassiter, et al, *Neglected Ethical Issues in Biobank Management: Results from a U.S. Study*, 9 Life Sci. Soc'y Pol'y 1 (December 2013) (estimating about 800 biobanks in the United States); Hana Odeh et al, *The Biobank Economic Modeling Tool (BEMT): Online Financial Planning to Facilitate Biobank Sustainability*, 13 Biopreserv. Biobank. 421 (2015) (estimating thousands of biobanks); Bryan Keogh, *European Biobanks Forge Cross-Border Ties*, 103 J. Nat. Cancer Inst. 1429 (2011) (estimating tens of thousands of biobanks globally).

type.[59] Myriad again provides the poster child for such private data siloing.[60] Nonprofit biobanks are still driven by incentives that may encourage fragmentation, including access to grant funding, prestige, or focus on particular diseases, or national mandates. The biobank operated by Partners Healthcare in Cambridge, Massachusetts, for instance, provides samples only to about 6,000 researchers affiliated with Partners, for approximately $20 each, potentially raising concerns about access both in terms of cost and in terms of who can reach the resources.[61] In addition to the problem of mixed incentives for connecting and sharing data, biobanks that do wish to share data face other obstacles, including data regulations such as the HIPAA Privacy Rule in the United States and the General Data Protection Regulation in the European Union; the latter will impact both intra-EU data use and sharing of data between US and EU biobanks.[62]

Some biobanks are already addressing fragmentation concerns, such as the set of European biobanks that have adopted the MIABIS framework to help facilitate connection.[63] And other scholars are convening and working to address questions of fragmentation, connection, interoperability, and access.[64] But substantial challenges remain as many biobanks globally keep their resources proprietary and thus leave potential innovations in translational medicine, or other areas, on the table.

3 CONCLUSION

Biobanks hold tremendous possibilities to serve as innovation infrastructure for translational medicine. They generate and store both biomedical samples and data, and these resources can be used to help bring innovation from the basic research laboratory into clinical practice. Among other challenges,

---

[59] *See*, *e.g.*, Arti K. Rai, *Risk Regulation and Innovation: The Case of Rights-Encumbered Biomedical Data Silos*, 92 Notre Dame L. Rev. 1641 (2017).

[60] *See*, *e.g.*, Robert Cook-Deegan et al, *The Dangers of Diagnostic Monopolies*, 458 Nature 405 (2009); *but see* Dan L. Burk, *Patents as Data Aggregators in Personalized Medicine,* 21 B.U. J. Sci. & Tech. L. 233 (2015) (describing how Myriad used its patents to aggregate—and partially defragment—large amounts of data about breast cancer).

[61] Beth Daley & Ellen Cranley, *The Rise of Bio-Rights: Patients Demand Cash for DNA Samples*, The Eye (October 10, 2016), https://eye.necir.org/2016/10/10/rise-bio-rights-patients-demand-control-get-cash-dna-samples/.

[62] See Eisenberg & Price, *supra* note 6.

[63] Roxana Merino-Martinez et al, *Toward Global Biobank Integration by Implementation of the Minimum Information about Biobank Data Sharing (MIABIS 2.0 Core)*, 14 Biopreservation Biobanking 298 (2016).

[64] *See*, *e.g.*, Liddell & Liddicoat, *supra* note 42.

fragmentation of data reduces the ability of biobanks to play this role; in many instances, data are proprietarily maintained in biobank-specific silos. Biobanks should consider not only that they provide infrastructure for their data, but also that those data themselves can serve as infrastructure for forward-looking innovation. With better connection, interoperable data, and technical standards, and broad accessibility to researchers, biobanks and networks of biobanks can play a larger role in facilitating translational medicine.