

“A Double-Edged Sword”: A Brief History of Genomic Data Governance and Genetic Researcher Perspectives on Data Sharing

Kayte Spector-Bagdady,¹ Kerry A. Ryan,¹ Amy L. McGuire,² Chris D. Krenz,³
M. Grace Trinidad,⁴ Kaitlyn Jaffe,⁵ Amanda Greene,¹ J. Denard Thomas,¹
Madison Kent,⁶ Stephanie Morain,⁷ David Wilborn,¹ and J. Scott Roberts¹

1. UNIVERSITY OF MICHIGAN, ANN ARBOR, MICHIGAN, USA. 2. BAYLOR COLLEGE OF MEDICINE, HOUSTON, TEXAS, USA. 3. BOSTON UNIVERSITY, BOSTON, MASSACHUSETTS, USA. 4. IMAGING DATA COMMONS, NEEDHAM, MA, USA. 5. UNIVERSITY OF MASSACHUSETTS AMHERST, AMHERST, MASSACHUSETTS, USA. 6. UNIVERSITY OF CALIFORNIA, LOS ANGELES, LOS ANGELES, CALIFORNIA, USA. 7. JOHNS HOPKINS UNIVERSITY, BALTIMORE, MARYLAND, USA.

Keywords: Genetics, Data Sharing, Genetic Testing, National Institutes Of Health

Abstract: As the federal government continues to expand upon and improve its data sharing policies over the past 20 years, complex challenges remain. Our interviews with U.S. academic genetic researchers (n=23) found that the burden, translation, industry limitations, and consent structure of data sharing remain major governance challenges.

In August 2022, the Biden White House’s Office of Science and Technology Policy (OSTP) released a memorandum on *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research* (2022

OSTP Memo).¹ Its major aims include requiring that federally supported research, along with data of sufficient quality to validate and replicate the findings, be made available to the public without embargo. The 2022 OSTP Memo represents the most recent step in federal data sharing efforts over the past 20 years, including those specific to genomic data sharing.² Since the last OSTP Memo on this topic in 2013, all 20 federal departments and agencies covered within OSTP’s scope have implemented data sharing policies. These policies have enabled access to over 2.4 million federally supported publications and an additional 5.7 million articles in the sciences generally.³

Despite these achievements, many of the problems these federal policies set out to fix remain. Among these challenges are: (1) clarifying who should bear the burden of sharing data; (2) translating shared data into scientific advancements; (3) elucidating how federal

Kayte Spector-Bagdady, J.D., M.Be., is an assistant professor of obstetrics and gynecology and interim co-director of the Center for Bioethics & Social Sciences in Medicine at the University of Michigan Medical School in Ann Arbor, MI. **Kerry A. Ryan, M.A.,** is a Research Area Specialist Senior at the Center for Bioethics and Social Sciences in Medicine at the University of Michigan Medical School in Ann Arbor, MI. **Amy L. McGuire, J.D., Ph.D.,** is the Leon Jaworski Professor of Biomedical Ethics and director of the Center for Medical Ethics and Health Policy at Baylor College of Medicine in Houston, TX. **Chris D. Krenz, B.A.,** is a former Research Associate at the Center for Bioethics and Social Sciences in Medicine at the University of Michigan Medical School in Ann Arbor, MI, and is currently a Master’s student in Electrical and Computer Engineering at Boston University in Boston, MA. **M. Grace Trinidad, Ph.D., M.P.H., M.S.,** is the Research Director of the Future of Trust agenda program at Imaging Data Commons. **Kaitlyn Jaffe, Ph.D.,** is an assistant professor of community health education at the University of Massachusetts Amherst in Amherst, MA. **Amanda K. Greene, Ph.D.,** is a Research Specialist at the Center for Bioethics and Social Sciences in Medicine at the University of Michigan Medical School in Ann Arbor, MI. **J. Denard Thomas, Ph.D.,** is a Research Area Specialist at the Center for Bioethics and Social Sciences in Medicine at the University of Michigan Medical School in Ann Arbor, MI. **Madison Kent, B.A.,** is an M.D. candidate at the David Geffen School of Medicine at the University of California, Los Angeles in Los Angeles, CA. **Stephanie Morain, Ph.D., M.P.H.** is a core faculty member at the Berman Institute of Bioethics, and an associate professor in the Department of Health Policy and Management at the Bloomberg School of Public Health. **David S. Wilborn** is an undergraduate research assistant at Michigan Medicine. **J. Scott Roberts, Ph.D.,** is Professor of Health Behavior & Health Education at the University of Michigan School of Public Health in Ann Arbor, MI.

policies intersect with private interests (e.g., journals, industry co-funders, or commercially generated data); and (4) balancing the autonomy interests of those who contribute data (including patients, research participants, and commercial consumers) with the public beneficence attendant to advancing science.

Due to the need to combine large amounts of data globally to support comprehensive advances across genomic variance, health behaviors, and health outcomes, the governance of genomic data sharing was largely where these types of policies began — and the field of genetics remains on the cutting edge of the debate regarding ongoing challenges. Therefore, while the U.S. government continues to focus on implementation of the 2022 OSTP Memo, and the National Institutes of Health (NIH) is concurrently updating

principles for the HGP regarding the sharing of genomic data, known as the “Bermuda Principles.”⁴ These principles mandated that sequencing data should be “freely available and in the public domain” to enable research, development, and the betterment of society.⁵ NHGRI then expanded the scope of these principles from the HGP to all its funded large-scale researchers, which evolved several times through 2003.⁶

2003 NIH Policy

In 2003, the NIH adopted a federal data sharing policy across all institutes and centers, called the *NIH Data Sharing Policy and Implementation Guidance* (2003 NIH Guidance). It required that investigators asking for \$500,000 or more in direct costs per year have a “plan for sharing final research data for research pur-

In this article, we provide a background of the major U.S. federal government data sharing policies over the past twenty years, present the results of our qualitative study, and discuss areas for continued improvement for federal governance and support of research.

its most recent 2014 genomic data guidance, it is critical to better understand the goals and challenges of those expected to both benefit from and contribute to these shared data resources. To this end, in the spring and summer of 2020, we conducted semi-structured interviews with U.S. academic genetic researchers. We explored perceived benefits and burdens, industry interests, and autonomy considerations related to data sharing and using shared data resources. In this article, we provide a background of the major U.S. federal government data sharing policies over the past twenty years, present the results of our qualitative study, and discuss areas for continued improvement for federal governance and support of research.

Background

1997 Bermuda Principles

U.S. science funding agencies began in the 1980s to think comprehensively about data sharing from funded research. The Human Genome Project (HGP), with the goal of generating the first sequence of the human genome, was launched in 1990. United States participants were funded by the U.S. Department of Energy and the NIH Office for Human Genome Research (later named the National Human Genome Research Institute (NHGRI)). Six years later, 50 members of the HGP gathered to adopt the first major set of prin-

poses, or state why data sharing is not possible.” The sharing had to “occur in a timely fashion” (generally defined as “no later than the acceptance for publication of the main findings from the final dataset”) and contain information necessary to “document, support, and validate” research findings. Such data also had to include relevant information about methods, codes, variables, etc. needed to “prevent misuse, misinterpretation, and confusion.”⁷

The policy also explicitly recognized that “the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort.” It specifically allowed investigators to benefit from “first and continuing use but not from prolonged exclusive use” of the data they generated. The 2003 NIH Guidance was also particularly concerned about the generation and analysis of data that had been “co-funded” by private industry. It recognized “the need to protect patentable and other proprietary data,” if those limitations were disclosed in the original grant proposal’s data sharing plan. The NIH also recognized the rights of contributors to privacy protections. However, it also recommended that promises to contributors that their data would not be shared as part of the informed consent or disclosure process “should not be made routinely and without adequate justification.”⁸

2008 NIH GWAS Policy

In 2008, the NIH created its own genomic data sharing policy common across institutes and centers (i.e. not just limited to NHGRI). The *Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies* (2008 NIH GWAS Policy) created a centralized data repository (the database of Genotypes and Phenotypes (dbGaP)), protected data contributors by ensuring that data sharing did not run contrary to the terms of the informed consent, and set standards for publication and intellectual property rights for all NIH-funded research that included GWAS. The policy required the sharing of protocols, instruments, variables, and supporting documentation, and “strongly encourage[d]” the sharing of curated phenotypic and genomic data within dbGaP. It also granted awardees a period of up to 12-months publication exclusivity from the shared dataset (others were allowed to analyze the data, but not submit findings to a journal during that time).⁹ Under this policy, over 2,200 investigators accessed 304 studies and produced over 900 publications.¹⁰

2013 OSTP Memo

In 2013, under the Obama Administration, the White House’s OSTP released its own *Increasing Access to the Results of Federally Funded Scientific Research Memorandum* (2013 OSTP Memo) to set one of the first federal data sharing standards, again increasing the coverage of data sharing requirements to now include many federal departments and agencies who fund research. The goal of the 2013 OSTP Memo was to “maximize the impact and accountability” of federal investment in research to “accelerate scientific breakthroughs and innovation.” It included the 20 federal departments and agencies with over \$100 million in annual research and development expenditures in its scope.¹¹

The 2013 OSTP Memo set the same 12-month post-publication embargo period for making all research papers “directly arising from federal funding” publicly available as the NIH had in 2003. It also required the sharing of data “commonly accepted in the scientific community as necessary” to validate the findings described therein. In addition, the 2013 OSTP Memo recognized the importance of balancing the ambitious goals of data sharing with “associated costs and administrative burden.” The memo specifically emphasized an interest in not adversely affecting opportunities for non-federally funded researchers, although it did not offer clear guidance regarding how to do so.¹²

Much like the 2003 NIH Policy, the 2013 OSTP Memo recognized proprietary interests to avoid “sig-

nificant negative impact on intellectual property rights, innovation, and U.S. competitiveness.” This time, with the addition of the article sharing requirement, it also recognized the interests of journals as discrete stakeholders. OSTP argued that “publishers provide valuable services, including the coordination of peer review, that are essential for ensuring the high quality and integrity of many scholarly publications.” It therefore required agency plans to have a strategy for “leveraging existing archives...and fostering public-private partnerships with scientific journals” as well as procedures to help prevent the “unauthorized mass redistribution of scholarly publications.” To maximize the impact of federal funding, it specifically encouraged public-private collaboration to maximize interoperability and creative reuse. In addition, the 2013 Memo noted the need for agencies to ensure that “confidentiality and personal privacy” of contributors were protected throughout.¹³

2014 NIH GDS Policy

The following year, the NIH replaced its 2007 NIH GWAS Policy with the *Genomic Data Sharing Policy* (2014 GDS Policy). This current policy applies if federal funding supports the “generation” of genomic data. While it did not alter OSTP’s required 12-month embargo for release of federally funded articles, it offered additional details to ensure “broad and responsible sharing” of large-scale genomic data. The 2014 GDS policy requires funded investigators to share genomic data, including the analytic code or tools necessary to interpret it, in an NIH-designated repository by the time of publication of their first related article.¹⁴

Public comments to this proposed policy expressed concerns regarding the financial burden that such a detailed level of data sharing would place on investigators, emphasizing the related infrastructure needed for such data sharing and the reallocation of already limited resources away from primary research. In addition, critics pointed out that the timeline for sharing genomic data could limit researchers’ ability to “perform adequate quality control.” The NIH acknowledged the “significant effort to prepare the data for sharing,” but maintained that this burden was “warranted by the significant discoveries made possible through the secondary use of the data.”¹⁵

Notably, the 2014 GDS Policy requires investigators to request informed consent for future use and sharing of genomic data derived from cell lines or clinical specimens collected after the effective date. The federal regulations, which set the requirements for human subjects research, Subpart A of which is

called the “Common Rule,” do not cover de-identified biospecimens and therefore do not require informed consent for de-identified specimen sharing.¹⁶ But the 2014 GDS Policy tightened this standard, arguing that “it is increasingly clear that participants expect to be asked for their permission to use and share their de-identified specimens for research,”¹⁷ even if those specimens are de-identified as defined by the HIPAA Privacy Rule (e.g., lacking name or address).¹⁸ This sets up a bifurcated system in which these additional protections do not apply to de-identified data, but *do* apply to the de-identified specimens from which those data are derived in the first place.

This federal justification for requiring informed consent for research with de-identified specimens and cell lines mirrors that which was used in 2015, when the U.S. Department of Health and Human Services released a Notice of Proposed Rulemaking to update the Common Rule.¹⁹ It too proposed that de-identified data remain outside the protections of Common Rule, but that the regulations should be changed to newly cover de-identified specimens; it even cited the same three underlying studies as the 2014 GDS Policy to support this claim.²⁰

That said, many commentators on the Notice of Proposed Rulemaking argued against the proposal to treat all biospecimens as inherently identifiable, due to concerns regarding making specimen research more expensive, less common, and restricting research productivity overall.²¹ The final revisions to the Common Rule therefore did not adopt this proposal writ large.²² The informed consent requirement for de-identified specimen research remains limited to federally-funded studies that generate genomic data. The only allowable exceptions to the informed consent requirement in the 2014 GDS Policy must be for “compelling scientific reasons.” Funded investigators are to request contributor consent for the “broadest possible sharing” but, if not, investigators are to submit data to controlled-access repositories.²³

2020 NIH Policy

A new *NIH Policy for Data Management and Sharing* (2020 NIH Policy) updates the 2003 NIH Policy in several important ways. These include broadening the scope of covered research, from that which cost \$500,000 per year, to “all research, funded or conducted in whole or in part by NIH, that results in the generation of scientific data.” In addition, while it maintains the requirement that data must be shared by the time of the first associated publication, it adds that even data that are not ultimately published must be shared by the end of the award period — whichever

comes first. It invokes a standard for sharing “quality” data, which includes both the ability to validate and *replicate* research findings whether or not those findings are ultimately published.²⁴

It also requires investigators to “maximize” the amount of data that can be shared (e.g., through the informed consent process), but acknowledges the potential “ethical, legal, or technical” factors that might limit such sharing. It encourages investigators to ensure that contributors are informed regarding what will happen with their data to respect their autonomy, and that factors that might impact sharing (e.g., limitations on consent for certain types of research) “travel” with the data to inform future users.²⁵ This policy became effective in January 2023 and includes a commitment to updating the 2014 NIH GDS Policy, as well.²⁶

2022 OSTP Memo

The most recent federal data-sharing memorandum, *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research Memorandum* (2022 OSTP Memo), was released in August 2022. Its goals include enhancing equity and trust in government-supported science, and it broadens the scope of federal departments and agencies that must develop their own data sharing policy from those with over \$100M in R&D funding to those with *any* funding.²⁷

Perhaps most notably, it responds to what it describes as “years of public feedback” that the 12-month embargo period was “inequitable” in that it limited “immediate access [to published articles] to only those able to pay for it or who have privileged access through libraries or other institutions.” The 2022 OSTP Memo therefore requires that all published articles resulting from federal funding (including funding held by co-authors) be made “freely available and publicly accessible” without journal embargo or delay.²⁸ OSTP also included highly stipulated guidance regarding the kinds of repositories in which investigators should deposit their data.²⁹ These recommend that repositories should provide free and easy access,³⁰ curation and quality assurance,³¹ common formatting,³² clear provenance,³³ and fidelity to consent.³⁴

In an attempt to move away from the “financial means and privileged access,” which, OSTP argued, are currently required to access cutting-edge scientific findings, the 2022 OSTP Memo cites values of “equal opportunity” in allowing “all Americans to benefit from the returns on our research and development investments without delay.” It delegates the National Science and Technology Council’s Subcommittee on Open Science to develop measures to additionally

reduce inequities for “individuals from underserved backgrounds and those who are early in their careers,” as well as reduce the burden of data sharing on funded researchers generally.³⁵

While the 2022 OSTP Memo also gives the Subcommittee on Open Science the task of coordinating engagement with stakeholders, “including but not limited to publishers...,” it lacks similar deferential language regarding the 2013 OSTP Memo’s concerns about publishers’ value to the research enterprise. It also adds new language regarding transparency surrounding the generation of federally funded scholarship, including “authorship, funding, affiliations, and development status” of the work.³⁶

Present Study

Before the 2022 OSTP Memo was released, we conducted semi-structured interviews with U.S. academic genetic researchers. While these interviews focused on genomic data specifically (i.e., the researchers were sampled via a PubMed publication of an article including genomic data), they discussed both genomic and other related phenotypic data. Previous data sharing policies have focused on data sharing with limited exploration of the related burden on funded researchers, a definition of industry partnership that no longer covers the complex scope of current data sharing partnerships, and somewhat contradictory stances on respecting contributor autonomy (e.g., discouraging participants from opting out of data sharing but also requiring consent for some specimen use). We therefore conducted this study to provide insights into the impact of federal data-sharing policies, with a focus on genomic data, through a qualitative exploration of perceived benefits and burdens of both sharing and using shared data resources, the translation of shared data into improved science, challenges with weighing industry interests, and considerations regarding informed consent under this dynamic governance landscape.

Materials And Methods

Recruitment

We identified prospective interviewees based on a PubMed review of 2017 – 2019 articles with at least one U.S. academic-affiliated corresponding author, which also indicated use of genomic data from at least one of the following types of genomic data stewards (i.e., entities that govern or oversee data resources): (1) A private steward (based on their inclusion in Research and Markets’ rank of direct-to-consumer (DTC) genetic testing companies, i.e., 23andMe, Ambry Genetics, Ancestry.com, Color Genomics, Gene

by Gene),³⁷ or (2) An academic, government, or consortia-related steward. We wanted to ensure that half of our sample used private stewards due to our specific interest in querying the under-explored relationship of the impact of private genomic data on research. The other half of our sample used non-private stewards, which ended up representing academic, government, and consortia-controlled data resources. We contacted the authors of approximately half of the identified articles – starting with those published most recently to aid in interviewee recollection and oversampling for female and Latino/Hispanic, African American or Black, or Asian researchers – via an email to the corresponding author (46% response rate). A more detailed description of recruitment is available in a previously published paper from these interviews.³⁸

Interviews and Analyses

We generated a semi-structured interview protocol based on a literature review of different attributes of genomic databases and solicited input from qualitative methods experts and genetic researchers to identify confusing or unclear phrasing prior to recruitment. We asked interviewees questions regarding employment, why they chose a specific data steward(s) to answer their research question (if they had the choice to begin with), contributor protections, data usage agreements, funding, data-sharing, and research outcomes (our interview guide is available as an appendix to a previous publication³⁹).

In our previous analysis, we focused on interviewees’ selection of database(s).⁴⁰ Here, we focus on researcher perspectives regarding sharing their genomic and related phenotypic data, as well as using data shared by others. While interview questions focused on the database(s) identified in the author’s PubMed publication result, we also asked them to compare this with their experiences using other databases.

We carried out each 30 to 60-minute interview via Zoom or telephone between March and July 2020 (KSB, CK, MK). Our male and female-identifying interviewees were non-Hispanic White and none of the interviewees conduct their own research with genomic data. We provided interviewees with a \$100 gift card following completion of the interview. We audio recorded and transcribed the interviews, reviewed the resulting transcripts for accuracy, and cleaned and de-identified them (CK). For the thematic analysis, we employed a method of iterative description, using grounded theory.⁴¹ We characterized themes common across interviews and captured individual variation. (KSB, KR, MGT, CK).

Our preliminary codebook was developed based on the structure of the interview guide, and then was iteratively edited after initial review and analysis of transcripts. All analysts concurred that thematic saturation was reached after 23 interviews. We then double-coded all transcripts (KR, MGT, CK) and met as a team to reconcile any discrepancies (KR, MGT, CK, KSB). We read through coded excerpts to identify relevant themes, which were then discussed with the entire team and consolidated into the final thematic analysis. This study was approved by the University of Michigan Institutional Review Board (HUM00175088), and each participant provided informed consent.

Results

Out of the 23 U.S. academic genetic researchers we interviewed, eleven used a private database in their reference article, and 12 used an academic, government, or consortia database. The majority of interviewees were female (n=13), non-Hispanic White researchers (n=14), with an average of 8.5 years at their current institution (see our previous publication for demographic tables⁴²). Nearly all compared different types of databases beyond the one for which they were sampled, leading to a discussion of 70 distinct databases (30 academic data stewards, 13 government, 11 private, 8 NGOs, and 8 via collaborations).

Theme 1: Sharing Data was Seen as A Burden Without Reward

A major challenge discussed in all the federal data sharing policies is who should carry the burden of data sharing, and how to limit its weight. Our interviewees described cleaning, preparing, and depositing data into authorized government repositories as laborious for investigators and their teams. One interviewee believed that this problem was particularly compounded at primary data collection sites:

...those investigators are sort of like ‘we hate actually being one of the funded sites because we make all the phenotypes and all genotypes available immediately, and we’re so busy collecting all the data that we don’t even have time to analyze it.’ ... So, [mandated data sharing is] sort of a double-edged sword...

Data sharing requires either the investigators take on the task themselves, “which is a huge undertaking,” or pay others to do it. But another interviewee described the problems of data sharing and cleaning even *if* the government provided funding for assistance (as it currently does). Data sharing is complex and requires a

baseline of expertise — but lacks attendant academic prestige. Thus, even when paid, the task was considered undesirable:

Keeping our labs motivated, keeping our post docs motivated, keeping them productive is hard enough and then having [to make] them go through some really cumbersome process to make their data available, which involves both bureaucratic work and work organizing and curating the data, which people don’t often see benefit from? So, yeah, I think it’s a lot of things that make [data sharing] challenging.

Not only was data sharing described as lacking academic prestige, but several interviewees also complained about the potential *loss* of academic opportunities in so doing. For example, one interviewee, discussing the current requirement of sharing project data (in effect since the 2003 NIH Memo), described the general hesitation that, if investigators share their data while still in the process of analysis for subsequent publications or grants, there could be another researcher who would “beat you, quote unquote, to the punch to find that new discovery within your own data.” Among researchers, this phenomenon is commonly referred to as being “scooped.”

In addition to receiving credit for a new discovery, this interviewee was particularly worried about securing additional grant funding if supporting preliminary data were already published by others: “I can be a good citizen, but how do you get a return on investment, right?” Although data sharing delays are supposed to be limited by the current federal requirements, interviewees also described how the lack of enforcement of those requirements compounded these issues as well as researcher uncertainty about the cost-benefit calculation of sharing data. One interviewee noted that current enforcement is “pretty bad in a lot of cases,” potentially unfairly compounding the burden of compliant researchers by enabling free riders who do not adhere to data sharing requirements.

As one interviewee summarized, “everyone needs to make this process easier” to enable investigators to share their data in the first place. The NIH puts “so much back on the researcher to make [data sharing] happen that I think it needs to be a little bit more centralized. Be *sure* it happens.”

Theme 2: Shared Data Often Lack the Quantity or Quality Necessary to Improve Science

As discussed above, the overarching goal of the federal data sharing policies is to improve science. But a

second theme of our interviews was that shared data sometimes lacks the quality to validate (required since 2003) and replicate (since 2020) research findings. The recent National Science and Technology Council's report on *Desirable characteristics of data repositories for federally funded research*, which came out two years after these interviews, includes the need for repositories in the future to provide "curation and quality assurance" to improve "accuracy and integrity"⁴³ as well as "clear provenance."⁴⁴ Demonstrating how far shared data resources will have to go to meet these standards, our interviewees described a landscape of shared data that sometimes lacks the quantity and/or quality necessary to meaningfully translate those data into advanced knowledge.

For example, while one interviewee admitted that complaining about the lack of necessary shared (in this case, phenotypic) data "would not make me popular...amongst my peers," several interviewees reported such challenges. They described a lack of related clinical, supporting, or methodological data necessary to validate or replicate published results. As one interviewee stated, shared data are:

...basically provided in such small scale without the necessary information that's needed to really do the robust research needed...Like the [National Cancer Institute] has mandated data sharing for clinical trials, but [other investigators] upload publicly available just a fraction of the data you would need [to conduct a meaningful secondary analysis].

Although the requirement to share methods information has been in place since at least 2003, another interviewee discussed the lack of access to the methods by which datasets were generated. This forced them to go back and read related papers to try to understand how shared data were generated "and whether you think that was valid or not." Conversely, a different interviewee described their challenges in *sharing* such methodological information — especially when working in a large consortium where each site had different IRB and consent requirements and, therefore, different methodological scope.

Others voiced concerns about the quality of data that *were* shared. One interviewee stated that they would be worried about using a publicly available dataset where curation "was not rigorously performed and reputable..." As they pointed out, "It's very easy to put a whole bunch of crap out there..."

Further, demonstrating the circular nature of challenges with sharing data and using data that have been

shared, one interviewee observed that concerns about being "scooped" made them "feel like researchers sit on that data for a really long time because they want to get as much as they can from their labs before they share it," which in turn led to data being dated "by the time it gets released to everyone else."

Theme 3: Private Interests Can Limit the Amount of Data Funded Investigators Share

"Free and easy access" is another component of the new NSTC's desirable repository characteristics.⁴⁵ Our previous analysis indicated that the concept of "easy access" was most closely aligned with the use of private data stewards.⁴⁶ Thus, while the 2003 NIH Policy focused mostly on private "co-funding," and more recent policies have discussed industry interests in terms of publishers, our interviews focused on another component of public-private collaboration: funded researchers using privately held genomic data for their work.

About half of our interviewees described their experiences working with private data stewards. We have previously described the benefits interviewees perceived in working with private stewards,⁴⁷ but interviewees also discussed why private stewards wanted to work with *them*, as academic researchers, as well. Perceived advantages for private data stewards to collaborate included co-authorship, learning new methods, publicity, and the ability to attract new customers for direct-to-consumer genetic testing products. Altruism was also described as a driving motivation. Data stewards, including private ones, were seen as "happy to see that their data is used for interesting scientific questions."

One interviewee discussed the potential value for private data stewards of dataset validation when analyses relying on their data are published in reputable journals. For example, one interviewee said that when they published, the private data steward they worked with linked to their article on their website: "So that at least it looks like they are working with name brand institutions when other researchers are looking at the website to see whether they should work with them." Some private datasets rely on self-reported (as opposed to clinician or researcher-captured) phenotypic information, which has been criticized as potentially lacking validity.⁴⁸ But, one interviewee argued, publishing with this kind of data comes with it *de facto* validation of the underlying dataset itself:

...it legitimizes [the private data steward] as a company, makes them look better in their research...they get their genetic insights followed

up on and they prove that the way they collect data is valuable, mainly by the self-report, and maybe that helps them build a case for then selling the data to various drug development companies.

In addition, while the 2013 OSTP Memo specifically encouraged public-private collaboration to “maximize interoperability and creative reuse as well as the impact of federal funding,”⁴⁹ our work told a different story. Despite perceived benefits to private stewards of sharing data with academic researchers, several interviewees also spoke of challenges with intellectual property rights in private data which limited their sharing and utility. They specifically described their experiences with private data stewards who would not let them share proprietary data with either journals or government databases:

...this became a really important roadblock for us in terms of publishing the paper, because basically, the journal said, ‘Your paper’s interesting. We would love to see a revision. But you need to make the data available.’ And [the private data steward] said, ‘Well, we can’t do that.’

This interviewee found this experience particularly frustrating because they believed that this steward had not been candid regarding what the data sharing restrictions would be in advance, and the journal ultimately rejected the paper because they could not deposit the data “in dbGaP or something like that.”

Another interviewee said that it *was* “totally public” that this same data steward restricts external investigators to only publishing up to 10,000 single nucleotide polymorphism-level results per paper. But they took issue with its reported justification for this policy as protecting the privacy of participants:

We have other ways to protect against re-identifiability of participants that, I think, make those concerns irrelevant. For example, we round the summary statistics that we make publicly available to five decimal places. We don’t give the actual real frequencies in our data, we instead posted 1000 Genomes [Project] allele frequencies. I think those precautions eliminate any concern about re-identifiability...

This interviewee agreed with the previous one that such intellectual property stipulations ultimately limited the usefulness of sharing the results.

Last, an interviewee discussed the use of other federal funding linked to privately held genomic data: that of the Centers for Medicare & Medicaid Services. When patients receive clinical genomic testing, often the generation and analysis of those data are sent to private testing companies. This can defray clinical costs for hospitals in that it centralizes and externalizes the expensive process of genomic analysis. But those resultant data are then generally also considered the property of the private company that generated them — even if the patient used federally funded insurance to pay for it:

...our government is effectively paying for these tests to be done, but yet they have no obligation to deposit that data into publicly available resources for us to use.... I mean, there’s literally hundreds upon hundreds of thousands — if not millions — of patients who might have gotten some form of genomic testing, and that data is completely unavailable to us, even though the government paid, basically, for it.

Theme 4: Tensions Exist Between Broad Data Sharing and Contributor Consent

Our last theme surrounds the tension between data sharing and transparency with, and informed consent from, contributors. Federal data sharing policies encourage investigators to “maximize” data usage through the informed consent process. If there are exceptions to this maximal sharing, annotations for appropriate use are supposed to adhere and travel with data to limit future uses. But two interviewees discussed not actually knowing the institutional review board (IRB) rules for using secondary data to begin with: “...we just kind of had to make up everything as we went along.” Another acknowledged that they “don’t know what current guidelines or practices are actually for research use” but that they “always did wonder in the back of my head” whether patients knew that other researchers had access to identified information. A different interviewee pointed out the need for ongoing education because, even when they had informed consent discussions with their own contributors, “people sometimes would ask me: am I going to clone them? Like sit around cloning random people?!”

Others talked about concerns that the appropriate kinds of informed consent were not secured for banked data or specimens, and/or whether contributors understood what it meant in the first place. One interviewee stated that they “suspect that the people never really consented to giving the data” that were collected decades ago. In fact, only three interviewees

knew the informed consent status of the contributors in the genomic data used for their article — and all three only know in retrospect because the journal required them to disclose it.

Taken as a whole, these IRB and informed consent concerns could critically impact the ability of investigators to effectively use shared data resources, par-

can hire to review their research proposals, as problematic because contract review might not have the same quality of oversight. As opposed to academic IRBs, contract IRBs face pressure to be “in favor of the company’s wishes.” In terms of informed consent, they were also worried whether contributors to private databases realize “that their data could be sold to drug

While the federal government continues to iteratively design and implement data sharing policies for funded research, many challenges remain. The goal of improving accessibility and impact is laudable, but our study demonstrates that sharing data is seen as a laborious burden without academic reward, shared data often lack the quantity or quality necessary for translation into improved science, private interests can limit data sharing and usefulness, and tensions remain between contributor autonomy and the advancement of science.

ticularly when trying to combine different datasets — and the limitations of contributor comprehension of information even when full informed consent is offered. One interviewee therefore described informed consent status as:

...actually one of the big barriers to accessing data...there are a lot of datasets that we might have used, but the consent was actually more narrow, or precluded us actually considering or using that dataset.

Concerningly, this interviewee even speculated that some limitations on consent were intentionally drawn to avoid some of the burdens of data sharing described above:

There’s a balance between protecting individual study participants and data sharing. I think some scientists may act in bad faith and may tailor consents in ways that their data ends up not being able to be available, even though they can publish papers in journal and publish findings.

They went on to point out that the same might be true for IRB approval because future data sharing is also “all driven locally, right by your local IRB — but it’s also driven a bit by what you asked for as an investigator.”

Another interviewee specifically brought together concerns regarding IRB review and consent with private data stewards. The interviewee described the system of “contract IRBs,” which private companies

companies who are developing certain medications and then making money off those medications.”

Discussion

While the federal government continues to iteratively design and implement data sharing policies for funded research, many challenges remain. The goal of improving accessibility and impact is laudable, but our study demonstrates that sharing data is seen as a laborious burden without academic reward, shared data often lack the quantity or quality necessary for translation into improved science, private interests can limit data sharing and usefulness, and tensions remain between contributor autonomy and the advancement of science.

Importantly, the challenges our interviewees described are, by and large, neither novel nor due to rapidly changing technologies. They are the same challenges that the federal government has been grappling with for more than two decades. Our findings, rooted in the context of iterative data sharing policies, underscore important considerations for federal departments and agencies that are crafting data sharing policies in response to the 2022 OSTP Memo.

First, many interviewees bemoaned the time-consuming process of preparing and depositing data into authorized government repositories. Despite federal reassurance to researchers that related costs can be included in grant budgets, our interviewees highlighted the remaining tension that data sharing tasks require a high level of technical aptitude. These tasks are often assigned to post-doctoral fellows and other

junior researchers who have the technical ability to do the work but lack the attendant academic prestige or production of academic deliverables that will further their careers. In addition, fellows and junior researchers often transition institutions in short order, so incoming trainees often must learn the process from scratch — taking even more time away from publication-producing research. Critically, our interviewees emphasized that financial cost (which is reimbursable) is not as valuable to them as *time* (which is not). Moreover, data sharing even presents the possibility of academic vulnerability via the risk of getting “scooped.”

These findings contrast with how the same interviewees, discussed in our previous paper, described *using* shared data resources to avoid the time-consuming and expensive process of generating their own data.⁵⁰ The tension is cyclical: researchers can save time and money using previously generated data, but perceive time and money as being wasted when asked to share it themselves. The underlying concern seems to be that of the “free-rider,” researchers who access the benefit of common resources without contributing themselves. If researchers share their data, they want to be reassured that others’ data will also be there for them to use — a problem that several ultimately blamed on a lack of sufficient enforcement of data sharing quality and standards by the government.

Beyond the findings of our study, it is worth noting the additional burden introduced by the 2022 OSTP Memo includes the immediate release of articles resulting from federal funding (including that held by co-authors). While journals had generally accepted the previous 2013 policy of a 12-month embargo for federally funded research without additional publication charges, the 2022 Memo lacked the deferential language to publishers of its 2013 counterpart. The immediate release of the article and data upon publication will affect journals’ business models more substantially and may lead to expanded publication fees, even if the submitted article was not supported by federal funding.⁵¹

While the 2022 OSTP Memo states that funded researchers may include open access fees in their budget proposals,⁵² questions remain regarding whether this will limit the flexibility of scientific discovery. For example, needing to precisely estimate number of publications and related study costs years in advance, before even starting the work, will be challenging for prospective budget requests. In addition, substantial publication fees for research with a limited budget can disincentivize researchers from publishing all their findings and negative findings in particular — both of which are critical to informing the field and avoid-

ing publication bias. High open-access fees could also affect collaboration among researchers. One could envision a non-federally funded research team declining the contributions of an author who receives federal funding so as not to put them in the position of having to pay for immediate release. Or researchers might publish the minimum number of articles they feel is necessary with a citation to their federal funding, and others without the funding citation to avoid fees. It will be interesting to see whether this new rule will result in a net gain of the amount of work that cites federal funding.

A second, and related, theme of our interviews is that shared data sometimes lacks the quality to validate and replicate research findings. While much federal time and money has been devoted to encouraging and requiring mass data sharing, there is a dearth of empirical validation of the ability of those data to be translated into advanced science.⁵³ As the federal government continues to invest federal time and resources — as well as the time and resources of federally funded researchers — empirical validation necessary to support an actual cost/benefit analysis of resources is critical.

Our interviewees also described a high motivation of private data stewards to collaborate with academic researchers to validate and publicize their product. We have previously found that the number of academic publications using private genomic data has increased over time. In addition, almost half of publications from 2011-17 using sampled private genomic databases also cited at least some NIH funding for the research.⁵⁴ Private data stewards can then profit from selling access to these validated databanks to other industry players, as illustrated by the recent \$300M agreement between GlaxoSmithKline and 23andMe.⁵⁵

But the academic-private interplay is not quite so clear-cut. Interviewees described not being able to share privately generated genomic data due to intellectual property concerns and contractual limitations. While this is certainly understandable from a business perspective — genomic data are an asset — the role of federal funding in building this asset remains under explored.⁵⁶ The 2014 GDS Policy importantly limits the scope of its applicability to funding used in the “generation” of genomic data, but it is unclear how the broadened scope of the 2022 OSTP Memo will change that standard.

This leads to a complex balance between the impact of federal funds on data sharing. As one interviewee pointed out, genomic data that are privately held may have been generated via Centers for Medicare & Medicaid Services clinical funding in the first place. And,

as Alexis Walker found in her recent qualitative exploration of employees of private sector genomics, the vast amount of industry IP is actually “developed in academic labs ...funded by the taxpayer.” One of her interviewees therefore “found it a bit egregious” that industry is then allowed to take that intellectual property, market it, and sell it back to patients at “massive margins.”⁵⁷ If federal funding can be used to analyze genomic data that cannot ultimately be shared, and in so doing add value to the data as a business asset, there are potentially large gaps preventing the government from maximizing on its investment in such public-private partnerships — a specifically stated goal.

Our interviews also highlighted a tension between the federal push for data sharing and protections for transparency and contributor consent. Our interviewees struggled to convey what kind of informed consent, if any, was provided for the information they used in established databanks. Only interviewees who were required to report it to their journal knew. This finding is consistent with our previous research which found that the type of contributor consent is not disclosed in academic papers using privately held genomic data almost half the time.⁵⁸ One interviewee voiced the concern that investigators that share data might even weaponize consent requirements to intentionally disallow themselves from sharing data in the future. This would both avoid the perceived burden of data sharing as well as the risk of being scooped. A lack of information regarding type of consent for shared data generally limits researcher ability to adhere to such standards, as well as government enforcement of their requirements. While the 2022 OSTP Memo lacks discussion of the type of consent necessary in its new language requiring transparency, implementing departments and agencies should consider this key component of disclosure and enforcement.

In addition, the current 2014 GDS Policy, by applying protections to de-identified biospecimens but not de-identified data, *de facto* assumes that contributors are more concerned about protections for the research use of their specimens versus data.⁵⁹ This was an argument also made by the 2015 Notice of Proposed Rulemaking for the Common Rule (but was not included in the final 2018 revision due to public response).⁶⁰ Both pieces even cite the same three articles to support this claim.⁶¹ None of the articles, however, actually do so. The first, Kaufman et al., elicits participants’ willingness to participate in a biobank — but did not actually compare participants’ attitudes regarding specimens *against* those regarding data.⁶² Vermeulen et al. again only queried (Dutch) patients about consent preferences for specimens,⁶³ and Trinidad et al. was norma-

tive and made no such comparative argument.⁶⁴ In fact, when we recently surveyed a national sample of over 2,000 patients, as opposed to finding that respondents were more likely to want notice regarding the future use of their specimens, we found that respondents were more likely to want notice regarding use of their *health information*.⁶⁵ Given the burden this biospecimen exceptionalism additionally poses for researchers, any new GDS policy potentially should re-consider this bifurcated requirement.

It is important to note that the findings reported in this research represent only a snapshot of the experiences of a relatively small sample of genetic researchers sharing and using shared data resources. Further research is necessary to generalize their experiences, such as surveys across a wider population, and an assessment of the relationship (if any) between researcher demographics, professional status, type of work, and experience. These interviews are a valuable step in this process.

Conclusion

As the federal government continues to expand upon and improve its data sharing policies over the past 20 years, complex challenges remain. Our findings demonstrate that the burden, translation, industry limitations, and consent structure of data sharing remain an issue. Thus, while the U.S. government continues to focus on this important work, it is critical that implementing departments and agencies better understand the goals and challenges of genetic researchers expected to benefit from, and contribute to, these broadened shared data resources.

Acknowledgements

The authors would like to thank the 2022 Health Law, Policy, Bioethics, and Biotechnology Workshop and Prof. I. Glenn Cohen at Harvard Law School, and Prof. Brian J. Zikmund-Fisher for their feedback on a previous draft of this paper. This work was funded in part by the National Human Genome Research Institute (K01HG010496 and T32-HG010030), the National Center for Advancing Translational Sciences (UL1TR002240, R01TR004244), the National Institute of Mental Health (R01MH126937), and the National Cancer Institute (R01CA237118).

Data Availability

De-identified qualitative quotes, organized by theme rather than in full transcript presentation to further protect the identity of the interviewee, is available upon request.

Note

This study was approved by the University of Michigan Institutional Review Board (HUM00175088). The study data were de-identified. This study was performed in accordance with relevant guidelines and regulations, including those set forth in the Declaration of Helsinki. Informed consent was obtained from all subjects, and de-identified data were used for analysis and reporting. The authors declare no conflict of interest.

References

1. Office of Science and Technology Policy (OSTP), Memorandum for the heads of executive departments and agencies: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research (August 25, 2022), *available at* <www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf> (last visited June 12, 2024) (hereinafter *2022 OSTP Memo*).
2. National Institutes of Health (NIH), GDS Policy Overview, *available at* <<https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing/gds-policy-overview>> (last visited June 12, 2024) (hereinafter *2008 NIH GDS Policy*).
3. OSTP, Public Access Congressional Report (November 5, 2021), *available at* <https://www.whitehouse.gov/wp-content/uploads/2022/02/2021-Public-Access-Congressional-Report_OSTP.pdf> (last visited June 12, 2024).
4. J.L. Contreras, “U.S. Federal Genomic Data Release and Access Policies,” *Bioinformatics, Medical Informatics and the Law*, J. L. Contreras, A. J. Cuticchia and G. Kirsch, eds. (Edward Elgar: 2021).
5. Human Genome Project Information Archive, “Policies on Release of Human Genomic Sequence Data Bermuda—Quality Sequence,” *available at* <https://web.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml> (last visited June 12, 2024).
6. J.J. Arias, G. Pham-Kanter, E.G. Campbell, “The Growth and Gaps of Genetic Data Sharing Policies in the United States,” *Journal of Law and the Biosciences* 2 no. 1 (2014): 56-68.
7. *Id.*
8. *Id.*
9. 2008 NIH GDS Policy, *supra* note 2.
10. NIH, “NIH Issues Finalized Policy on Genomic Data Sharing” (August 27, 2014), *available at* <<https://www.genome.gov/news/news-release/NIH-issues-finalized-policy-on-genomic-data-sharing>> (last visited June 12, 2024).
11. OSTP, Memorandum for the Heads of Executive Departments and Agencies: Increasing Access to the Results of Federally Funded Scientific Research (February 22, 2013), *available at* <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf> (last visited June 12, 2024) (hereinafter *2013 OSTP Memo*).
12. *Id.*
13. *Id.*
14. NIH, GDS Policy Overview, *available at* <<https://sharing.nih.gov/genomic-data-sharing-policy/about-genomic-data-sharing/gds-policy-overview>> (last visited *available at*) (hereinafter *2014 GDS Policy*).
15. *Id.*
16. 45 C.F.R. § 46 (2018).
17. 2014 GDS Policy, *supra* note 10.
18. U.S. Department of Health and Human Services, “Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule,” *available at* <<https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html>> (last visited June 12, 2024).
19. Department of Homeland Security, et al., “Federal Policy for the Protection of Human Subjects,” *Federal Register* 80, no. 173 (2015): 53933-54061, at 53942 (“...a growing body of literature shows that in general people prefer to have the opportunity to consent (or refuse to consent) to research involving their own biological materials.”) (hereinafter *Federal Policy for Human Subjects*).
20. D.J. Kaufman, J. Murphy-Bollinger, J. Scott, K.L. Hudson, “Public Opinion about the Importance of Privacy in Biobank Research,” *American Journal of Human Genetics* 85, no. 5 (2009): 643-654; E. Vermeulen et al., “A Trial of Consent Procedures for Future Research with Clinically Derived Biological Samples,” *British Journal of Cancer* 101, no. 9 (2009): 1505-1512; and S.B. Trinidad et al., “Research Practice and Participant Preferences: The Growing Gulf,” *Science* 331, no. 6015 (2011): 287-288.
21. H.F. Lynch, B.E. Bierer, I.G. Cohen. “Confronting Biospecimen Exceptionalism in Proposed Revisions to the Common Rule,” *The Hastings Center Report* 46, no. 1 (2016):4-5; L.H. Glimcher, “How Not to End Cancer in our Lifetimes,” *The Wall Street Journal* (April 4, 2016).
22. Federal Policy for Human Subjects, *supra* note 18, at 7168. (“The final rule does not implement the proposed expansion of the definition of “human subject” to include all biospecimens regardless of identifiability. It is clear from the comments received that the public has significant and appropriate concern about both the need for obtaining consent before using such biospecimens for research, and the potential negative impacts of implementing that proposal on the ability to conduct research.”)
23. 2014 GDS Policy, *supra* note 10.
24. National Institutes of Health, Final NIH Policy for Data Management and Sharing, NOT-OD-21-013, *available at* <<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>> (last visited June 12, 2024) (hereinafter *2020 NIH Policy*).
25. *Id.*
26. L.A. Jorgenson, C.D. Wolinetz, F.S. Collins, “Incentivizing a New Culture of Data Stewardship: The NIH Policy for Data Management and Sharing,” *JAMA* 326, no. 22 (2021): 2259-2260. (“The agency [NIH] intends to take a fresh look at existing data sharing expectations, particularly the 2014 NIH Genomic Data Sharing (GDS) Policy, to simplify compliance while achieving important policy objectives in the least burdensome way. Science, technology, and society views have all evolved in the years since the GDS Policy was issued, and the NIH intends to engage the research community and the public on a number of critical issues relating to the future of the GDS policy.”)
27. 2022 OSTP Memo, *supra* note 1.
28. *Id.*
29. National Science and Technology Council (NSTC), “Desirable Characteristics of Data Repositories for Federally Funded Research” (May 2022), *available at* <<https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf>> (last visited June 12, 2024).
30. *Id.* at 4 (“The repository provides broad, equitable, and maximally open access to datasets and their metadata free of charge in a timely manner after submission, consistent with legal and policy requirements related to maintaining privacy and confidentiality, Tribal and national data sovereignty, and protection of sensitive data.”)
31. *Id.* (“The repository provides or facilitates expert curation and quality assurance to improve the accuracy and integrity of datasets and metadata.”)
32. *Id.*, at 5 (“The repository allows datasets and metadata to be accessed, downloaded, or exported from the repository in widely used, preferably non-proprietary, formats consistent with standards used in the disciplines the repository serves.”)
33. *Id.* (“The repository has mechanisms in place to record the origin, chain of custody, version control, and any other modifications to submitted datasets and metadata.”)
34. *Id.* at 6 (“The repository employs documented procedures to restrict data access and use to those that are consistent with participant consent (such as for use only within the context of research on a specific disease or condition) and changes in consent.”)
35. 2022 OSTP Memo, *supra* note 1.
36. *Id.*
37. Research and Markets, Global Consumer DNA (Genetic) Testing Market—Forecasts from 2018–2023, accessed April 7, 2022, *available at* <https://www.researchandmarkets.com/research/w4fsmm/global_928?w=5> (last visited June 12, 2024).

38. M.G. Trinidad et al. "Extremely Slow and Capricious: A Qualitative Exploration of Genetic Researcher Priorities in Selecting Shared Data Resources," *Genetics in Medicine* 25, no. 1 (2023): 115-124.
39. *Id.*
40. *Id.*
41. S. Thorne, S.R. Kirkham, and K. O'Flynn-Magee, "The Analytic Challenge in Interpretive Description," *International Journal of Qualitative Methods* 3, no. 1 (2004):1-11; S. Thorne, S.R. Kirkham, J. MacDonald-Emes, "Interpretive Description: A Noncategorical Qualitative Alternative for Developing Nursing Knowledge," *Research in Nursing And Health* 20, no. 2 (1997):169-177; S. Thorne, *Interpretive Description: Qualitative Research for Applied Practice*, Second Edition, (New York: Routledge, 2016).
42. Trinidad et al., *supra* note 38.
43. NSTC, *supra* note 29, at 4.
44. *Id.*, at 5.
45. *Id.*, at 4.
46. Trinidad et al., *supra* note 37 ("...interviewees generally stated that ease of access was the most important factor in their selection, describing the interrelated components of familiarity with the data steward and efficiency.... Efficiency of database access was another central component of "ease of access," and was often described in terms a lack of legal red tape or a high customer service focus (which seemed to favor private stewards)").
47. *Id.*
48. S. Wyatt, A. Harris, S. Adams, and S. E. Kelly, "Illness Online: Self-Reported Data and Questions of Trust in Medical and Social Research," *Theory, Culture & Society* 30, no. 4 (2013): 131-150.
49. 2013 OSTP Memo, *supra* note 11.
50. Trinidad et al., *supra* note 38.
51. S. D'Agostino, "Who'll Pay for Public Access to Federally Funded Research?" *Inside Higher Ed* (September 12, 2022), available at <<https://www.insidehighered.com/news/2022/09/12/wholl-pay-public-access-federally-funded-research>> (last visited June 12, 2024).
52. 2022 OSTP Memo, *supra* note 1 ("...federal agencies should allow researchers to include reasonable publication costs and costs associated with submission, curation, management of data, and special handling instructions as allowable expenses in all research budgets.).
53. S.R. Morain, et al., "Ethics Challenges in Sharing Data from Pragmatic Clinical Trials," *Clinical Trials* 19, no. 6 (2022): 681-689 ("...additional empirical research should evaluate the demand for [pragmatic clinical trial] data, to inform assessments about how to tailor data sharing requirements to best ensure that shared data will, in fact, yield societal benefits so as to justify the investment required to broadly share those data.).
54. K. Spector-Bagdady et al. "Genetic Data Partnerships: Academic Publications with Privately Owned or Generated Genetic Data," *Genetics in Medicine* 21, no. 12 (2019): 2827-2829.
55. GSK, "GSK and 23andme Sign Agreement to Leverage Genetic Insights for the Development Of Novel Medicine," (July 25, 2018), available at <<https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/>> (last visited June 12, 2024); 23andMe, "23andMe Announces Extension of GSK Collaboration and Update on Joint Immuno-oncology Program," (January 18, 2022), available at <<https://investors.23andme.com/news-releases/news-release-details/23andme-announces-extension-gsk-collaboration-and-update-joint>> (last visited June 12, 2024).
56. K. Spector-Bagdady, "Governing Secondary Research Use of Health Data and Specimens: The Inequitable Distribution of Regulatory Burden Between Federally Funded and Industry Research," *Journal of Law and the Biosciences* 8, no. 1(2021): lsab008.
57. A. Walker, "Diversity, Profit, Control: An Empirical Study of Industry Employees' Views on Ethics in Private Sector Genomics," *American Journal of Empirical Bioethics* 13, no. 3 (2022):166-178.
58. Spector-Bagdady et al., Genetic Data Partnerships, *supra* note 54.
59. 2014 GDS Policy, *supra* note 10.
60. Federal Policy for Human Subjects, *supra* note 19, at 53942.
61. D.J. Kaufman et al., *supra* note 20; E. Vermeulen et al. *supra* note 20; and S.B. Trinidad et al., *supra* note 20.
62. D.J. Kaufman et al., *supra* note 20.
63. E. Vermeulen et al. *supra* note 20.
64. S.B. Trinidad et al., *supra* note 20.
65. K. Spector-Bagdady et al., "Reported Interest in Notification Regarding Use of Health Information and Biospecimens," *Journal of the American Medical Association* 328, no. 5 (2022): 474-476.