


RESEARCH ARTICLE

The National Cancer Institute Cancer Moonshot Public Access and Data Sharing Policy—Initial assessment and implications

Tammy M. Frisby¹ and Jorge L. Contreras^{1,2,*} 

¹University of Utah S.J. Quinney College of Law, Salt Lake City, Utah, USA

²Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah, USA

*Corresponding author. Email: jorge.contreras@law.utah.edu

Received: 22 February 2020; Revised: 11 May 2020; Accepted: 27 May 2020

Keywords: Cancer Moonshot; data commons; data sharing; dbGaP; NCI; open access

Abbreviations: CC, creative commons; CEDR, Cancer Epidemiology Data Repository; dbGaP, Database of Genotypes and Phenotypes; DMS, Data Management and Sharing; GDC, Genomic Data Commons; HGP, Human Genome Project; NCI, National Cancer Institute; NHGRI, National Human Genome Research Institute; NIH, National Institutes of Health; OMB, Office of Management and Budget; PADS, Public Access and Data Sharing; TCIA, The Cancer Imaging Archive.

Abstract

Since 2013, federal research-funding agencies have been required to develop and implement broad data sharing policies. Yet agencies today continue to grapple with the mechanisms necessary to enable the sharing of a wide range of data types, from genomic and other -omics data to clinical and pharmacological data to survey and qualitative data. In 2016, the National Cancer Institute (NCI) launched the ambitious \$1.8 billion Cancer Moonshot Program, which included a new Public Access and Data Sharing (PADS) Policy applicable to funding applications submitted on or after October 1, 2017. The PADS Policy encourages the immediate public release of published research results and data and requires all Cancer Moonshot grant applicants to submit a PADS plan describing how they will meet these goals. We reviewed the PADS plans submitted with approximately half of all funded Cancer Moonshot grant applications in fiscal year 2018, and found that a majority did not address one or more elements required by the PADS Policy. Many such plans made no reference to the PADS Policy at all, and several referenced obsolete or outdated National Institutes of Health (NIH) policies instead. We believe that these omissions arose from a combination of insufficient education and outreach by NCI concerning its PADS Policy, both to potential grant applicants and among NCI's program staff and external grant reviewers. We recommend that other research funding agencies heed these findings as they develop and roll out new data sharing policies.

Policy Significance Statement

Federal research-funding agencies must take into account the practical challenges inherent in promulgating broad new data sharing policies and mechanisms. This study identifies several implementation issues that hampered the rollout of National Cancer Institute (NCI)'s Cancer Moonshot Public Access and Data Sharing (PADS) Policy. The lessons learned by NCI should be applied to other agencies that are developing new data sharing policies, particularly as NIH considers a new agency-wide data sharing policy that would apply across all NIH institutes and centers.

Introduction

In 2016, the National Cancer Institute (NCI), the largest research funding organization within the U.S. National Institutes of Health (NIH), launched the ambitious \$1.8 billion Cancer Moonshot Program to accelerate innovative cancer research and advance cancer therapies and treatments. One of the key elements of the Cancer Moonshot Program, and of all cancer research today, is the generation of data. As one commentator notes, “[u]nderstanding cancer is understanding cancer’s *information*” (Sherkow, 2018, p. 301). And in order to ensure that the value of this data is maximized, and to overcome any inherent reluctance by cancer researchers to share their data, the Cancer Moonshot Program expressly encourages the broad sharing of data among researchers and the public (Cancer Moonshot Blue Ribbon Panel Report, 2016; Mattioli, 2017; Sherkow, 2018).

Key to the Cancer Moonshot Program was a new Public Access and Data Sharing (PADS) Policy (National Cancer Institute (NCI), 2017a; Department of Health & Human Services (HHS), 2017). The PADS Policy is applicable to all Cancer Moonshot funding applications submitted on or after October 1, 2017. It encourages the immediate public release of published research results and data and requires all Cancer Moonshot grant applicants to submit a PADS plan describing how they will meet these goals.

The PADS Policy is ambitious in that it extends data sharing principles that originated in the realm of genomics to all aspects of cancer research, from cell biology and pathology to chemotherapy and surgery to patient communication and psychology. But despite this sweeping mandate, little detail was provided in the Policy regarding data sharing practices and practicalities, particularly in research areas outside of cancer genetics and genomics. Informally, we observed that numerous cancer researchers funded by the Cancer Moonshot Program were perplexed by the requirements of the PADS Policy.

We thus sought to gain a better understanding of how cancer researchers understood and sought to comply with the PADS Policy requirements. To that end, we analyzed the PADS plans submitted with approximately half of all funded Cancer Moonshot grant applications in fiscal year 2018. We found that a majority of these researcher-generated plans did not address one or more elements required by the PADS Policy. Many plans made no reference to the PADS Policy at all, and several referenced obsolete or outdated NIH policies instead.

In the remainder of this article, we discuss the background data sharing requirements and policies that led to the Cancer Moonshot PADS Policy, as well as the specific requirements of the PADS Policy itself. We then describe the study that we conducted of submitted PADS plans and our findings. We conclude with recommendations for improving education and outreach about the PADS Policy, both to potential grant applicants and among NCI’s program staff and external grant reviewers, and how our findings may be applicable to broader data sharing plans that are currently under development at NIH.

Background

Origin and development of NIH data sharing policies

In 1996, leaders of the international Human Genome Project (HGP) established a baseline of rapid, public sharing of data in the biomedical sciences (International Human Genome Sequencing Organization, 1996). While the HGP was concerned primarily with genomic sequence data, the scope of scientific data sharing has expanded in the decades since its completion. Beginning in 2006, the NIH and its National Human Genome Research Institute (NHGRI) led the development of data sharing policies for a range of research programs generating data beyond raw DNA sequences, including genetic variants, DNA functional elements, and phenotypic and clinical data (Contreras, 2011; Contreras and Knoppers, 2018). Alongside these efforts, NIH instituted a policy in 2008 requiring that all publications resulting

from NIH-funded research be made publicly available through the National Library of Medicine's PubMed Central database (National Institutes of Health (NIH), 2008).

In 2013, seeking to extend the benefits of scientific data sharing to the full range of nondefense federally funded research, the Office of Management and Budget (OMB) enacted a sweeping Open Data Policy that required federal agencies to make most government-funded data available in accessible, modifiable and searchable formats, while taking into account individual privacy, confidentiality and security considerations (Office of Management and Budget (OMB), 2013). Initiatives such as these are aimed at the eventual development of an integrated "medical information commons" that will enable broad data use and analysis in the diagnosis of disease, the search for cures and the improved understanding of human biology (Bollinger et al., 2019).

Numerous federal agencies, including NIH, responded to OMB's ambitious directive. In 2014, NIH revamped its policy for genomic data sharing (National Institutes of Health (NIH), 2014), and, in 2015, NIH began planning a more sweeping overhaul of its trans-institute data sharing program (National Institutes of Health (NIH), 2015). Most recently, in November, 2019, it released a draft Policy for Data Management and Sharing (DMS) for public comment (National Institutes of Health (NIH), 2019). Like the Cancer Moonshot PADS Policy, the draft DMS policy mandates that researchers seeking NIH funding submit a data sharing *plan* as part of their grant applications. This plan would be reviewed by the funding agency in evaluating the application, and compliance would be monitored during the course of the funded research program.

The NCI Cancer Moonshot PADS Policy

In 2016, NCI, launched the Cancer Moonshot Program. The PADS Policy was an integral part of the program and its announcement (Kaiser 2019a). The Policy addresses both open access to publications resulting from Cancer Moonshot-funded research as well as the release of data generated by that research. In the area of publication, the PADS Policy takes a more aggressive stance than the 2008 NIH Public Access Policy, recommending (but not mandating) that research papers be deposited with PubMed Central immediately upon publication (and preferably within four weeks after acceptance by a journal), rather than one year after publication. The PADS Policy also encourages publication of papers under a Creative Commons attribution license (CC BY 4.0) or via contribution to the public domain (CC 0).¹

With regard to data, the PADS Policy states that "to the extent feasible," data should be made publicly available simultaneously with the publication of research results. Unlike prior NIH data sharing policies that pertained primarily to genomic and related data, the PADS Policy defines data that must be shared as any "recorded factual material commonly accepted in the scientific community as necessary to document and support research findings in Publications" (National Cancer Institute (NCI), 2017a). This broad definition thus encompasses all forms of clinical, pharmacological, demographic, analytical, survey, and other data that might be collected or developed as part of a Cancer Moonshot project.

Because the scope of data encompassed by the PADS Policy is so broad, NCI offers applicants a degree of flexibility in deciding how their research results and data will be released. Thus, the only firm requirement of the PADS policy is that applicants for Cancer Moonshot funding submit to NCI a written plan (a PADS plan) that describes a proposed process for making publications and data "immediately and broadly available to the public" and, if such sharing is not possible, a justification for why it is not. NCI also states that its review of grant funding proposals "will give funding priority to those Applicants that submit an appropriate [PADS] [p]lan that ensures maximal sharing of [publications and data] arising from the award."

¹ Creative Commons is a nonprofit organization that has developed a suite of self-executing licensing agreements for a wide range of copyrightable works. Most Creative Commons licenses (designated as "CC") permit the free dissemination, reproduction and modification of licensed works. CC BY licenses require only that the user give attribution to the original source of the work. Contributing a work to the public domain via a "CC 0" license requires the author to relinquish all rights to control or profit from future uses of the work (Creative Commons, 2019).

The PADS Policy lists the following “general elements” that each PADS plan “should, at a minimum, address”:

Release of publications

1. How publications will be made immediately available to the public.
2. The methods through which the public including other researchers will locate and access the publication.
3. Any anticipated limitations to the immediate and broad release of publications with an associated justification.

Sharing of underlying primary data

1. The type of data that is expected to be generated by the research.
2. The data that will be shared.
3. Who will have access to the data.
4. The timing and the medium for immediate sharing of the data.
5. The methods through which the public including other researchers will locate and access the data.
6. Any anticipated limitations to the immediate and broad sharing of data with an associated justification.

So, while the PADS Policy describes desirable features of submitted PADS plans (e.g., public release of publications within four weeks after acceptance), these features are not mandatory requirements, but merely recommendations.

Methods

To assess the degree to which successful Cancer Moonshot applicants addressed the requirements of the PADS Policy in its first year of implementation, we identified all competing grant and cooperative agreement applications that were funded under the Cancer Moonshot Program in fiscal year 2018 ($n = 63$) through NCI’s website (NCI 2019) <https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/implementation>. These projects were funded under eight of the 12 different Research Initiatives recommended by the Cancer Moonshot Blue Ribbon Panel. The Adult Immunotherapy Network initiative funded the largest number of projects with 16. The remaining Research Initiatives represented were Drivers of Childhood Cancer (11 funded projects), the Human Tumor Atlas Network (11), Symptom Management (8), Pediatric Immunotherapy Network (6), Prevention and Early Detection (6), New Enabling Cancer Technologies (4), and Hereditary Cancers (1). The principal investigator (PI) identified in each funded Cancer Moonshot application was then contacted by email with a request for the data sharing plan included with the funding application submitted to NCI.²

We received responses from 35 awardees³ representing all eight of the initiatives that funded proposals in fiscal year 2018. To respect our anonymity agreements with the researchers who responded, we limit

² In theory, this information could have been requested from NCI through a public records request under the Freedom of Information Act (FOIA). For this study, we did not submit FOIA requests as our past experience with FOIA requests made to NIH for other projects demonstrated that this process often involves significant delays. Moreover, we did not request amendments to grant applications made in response to input from NCI staff. While the scope of our findings does not extend to the final plans that were the product of informal and undocumented interaction with NCI staff after grant submission, we feel that there is value in assessing the alignment between investigators’ initial plan submissions and NCI’s written policy statements. One important measure of the success of a data sharing policy is how well the policy language, standing on its own, guides investigators in their efforts to follow the requirements and aspirations of the policy statement. The clarity of policy language is critical as funding programs scale up to larger and larger applicant pools, as it is not practical to assume that individualized guidance by program officers can be offered efficiently, affordably or equitably by every funding agency to every applicant. Moreover, from a legal standpoint, an agency’s formal written policies are important measures of its compliance with regulatory and statutory rulemaking and due process requirements.

³ Despite our high (>50%) response rate, we recognize that the self-selection of our respondents could have affected our findings. However, we think it is unlikely that self-selection significantly biased our results toward either more or less conformity with the

our reporting on the response rate within each research initiative to observing that we received plans from about half of the funded projects in each initiative. No single research initiative was significantly overrepresented in our overall response rate of approximately 55%. One initiative, Drivers of Childhood Cancer, was an outlier to the otherwise consistent response rate across research initiatives. We received three PADS plans from this group, out of the eleven funded projects.

The 35 responses came from 27 unique research institutions. Twenty-three institutions were awarded one Cancer Moonshot grant, two were awarded two grants, and two were awarded three grants. Each plan that we reviewed was submitted by a different PI. We received a total of 33 unique PADS plans⁴ and coded each with respect to the manner in which it addressed, or did not address, the different elements described in the PADS Policy.

We coded plans as satisfying a composite element of a PADS plan, for example, by identifying how access to publications would be provided or data would be shared, when language within a plan was associated with that element. This categorization was clearest when plans used—often word-for-word—the guidance examples provided in the PADS Policy. But we also coded plans as satisfying required plan elements when there were specific descriptions of other relevant actions that supported the particular element. We coded plans as representing a complete PADS plan when the submitted text addressed the required plan elements for both sharing of publications and underlying data. The text of each plan was manually parsed sentence by sentence and compared by the authors against the language from the PADS Policy that describes required elements and provides guidance examples.

Results

Overall structure of submissions

Among the 33 submitted plans that we reviewed, only 17 had substantial provisions related to both publication and data sharing, which would qualify them as PADS plans. In 13 of these, the plan was labeled a “Public Access and Data Sharing Plan” or used headings or other language to identify itself as a PADS plan. Every plan that addressed publication access also addressed data sharing. There were 8 additional plans that did not contain a section on open access publication, but which did provide significant information about data sharing, including guidance examples derived from the PADS Policy. Another 8 plans did not appear to address the PADS Policy elements for either publications or data sharing.

Highlighting the overlap between older NIH policies and the new Cancer Moonshot PADS Policy, 9 of the plans that we classified as partially or fully addressing the PADS Policy elements included references to only non-Cancer Moonshot NIH policies. The most common policies referenced were the 2003 NIH Statement on Sharing Research Data (11) (NIH 2003) and the 1999 NIH Research Tools Policy (12) (NIH 1999), which appeared in five and six plans, respectively. NIH’s 2008 Public Access Policy (4) (NIH 2008) was mentioned by one plan in this group.

Publication sharing

Under the PADS Policy, applicants are required to identify the methods through which the public, including other researchers, will be able to locate and immediately access publications generated by the grant-funded research. Of the 33 funded FY18 PADS plans that we reviewed, 16 did not engage with the

PADS Policy standards, given that we did not identify our specific research questions at the time that investigators’ PADS plans were solicited and that there was a high degree of variability in the plans that we received. If there is self-selection bias in our results, we expect that would be the product of a higher response rate among investigators who thought they had been largely successful in following the PADS Policy and, therefore, were comfortable sharing their PADS plan with other researchers. If that was the case, it would make our findings more remarkable because the low rate of adherence to the PADS Policy requirements that we see here would be an overstatement of how well PADS plans from fiscal year 2018 conformed to the new Cancer Moonshot PADS Policy.

⁴ Two responses were submitted by Coordinating Centers, which indicated that PADS plans would be developed by the individual research projects for which they acted as coordinator.

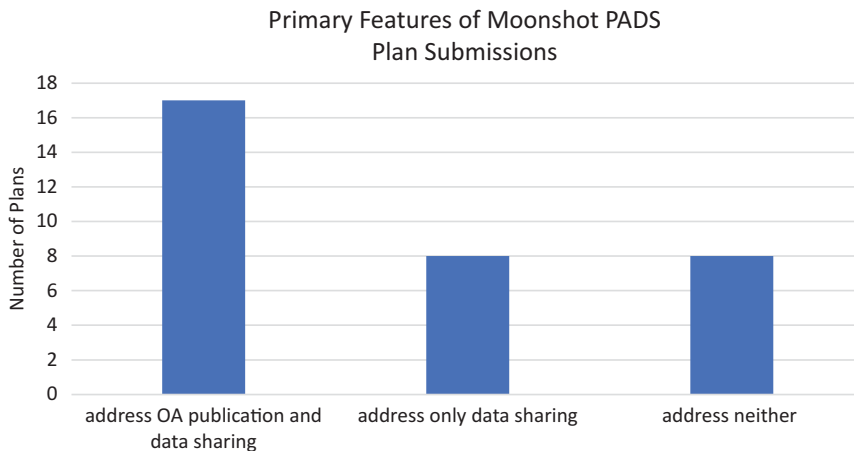


Figure 1. Primary features of PADS plan submissions.

PADS publication requirements. Two others mentioned publication without referencing the PADS Policy or mentioning the term “Moonshot.” One wrote of a preference for publishing in open access journals so long as the journal “commit[s] to rigorous peer review.” The other simply stated that the researchers would comply with the PubMed Central deposit requirement under the 2008 NIH Public Access Policy (which, as noted above, includes a 12-month delay in publication release that is inconsistent with the recommendations of the PADS Policy).

Among the 15 plans that more fully addressed the PADS Policy elements relating to publication, 11 inserted verbatim, or nearly so, the 3 relevant guidance examples from the PADS Policy, including the commitment to deposit publications in PubMed Central within 4 weeks after acceptance by a journal. Another stated that investigators would “consider” publishing in open access journals “whenever possible.” All 15 plans that addressed the PADS Policy publication elements included a statement that publications would be deposited in PubMed Central (as required since 2008).

Only three plans stated that the investigators would pay open access fees to ensure that publications would be made available on an open access basis. And only one of these noted that the grant proposal included a budget request for funds to support open access fees for accepted publications, charges that the PADS Policy explicitly designates as permissible budget items (we did not have access to the actual budget submissions or awards associated with these applications).

As noted above, the PADS Policy permits applicants to describe any proposed “limitations” on publication sharing. Nine of the 15 plans that included publication sharing provisions used language that limited unrestricted, immediate access to publications. Two of these expressly referred to limitations imposed by journal policies. For example, one stated, “After satisfying journal or society embargo periods, publications will be made publicly available immediately.” The same plan also committed to publish under Creative Commons or other public domain licensing “when not restricted by journal policy.” Two of the plans state that they will publish under an open access license “whenever possible,” and five use aspirational but noncommittal language about publication in open access journals. For example, these plans state that researchers “will seek” or “aim” or “consider” publication of results in open access journals. And one plan simply “pledge[d] support for open access publishing policies.”

NCI admitted that there was some confusion among grantees regarding the PADS Policy’s open access publication requirements, as well as resistance by publishers to allowing immediate release of published articles on PubMed Central (Fleming, 2019). To address these issues, NCI engaged in discussions with several major scientific publishers to explain the PADS Policy requirements, while at the same time reminding grantees that “It is imperative that all investigators supported by your Cancer Moonshot grant are aware of [the PADS] policy and are in compliance with the publication and data sharing plan of your award” (Fleming, 2019).

Data sharing

Twenty-five PADS plans contained specific provisions related to methods and means for sharing data. The PADS Policy asks investigators to specify the “medium” for data sharing. For this required plan element, NCI offers a guidance example that data will be made available through a repository such as Genomic Data Commons, the Database of Genotypes and Phenotypes (dbGaP), The Cancer Imaging Archive (TCIA), or a non-NIH repository that provides data access controls consistent with the PADS Policy. Twenty-two of the 25 plans addressing data sharing named at least one specific repository. The most common commitment, which was made by three plans, was to deposit data in dbGaP. In addition to two mentions each of the NCI Genomic Data Commons and the National Center for Biotechnology Information Gene Expression Omnibus, 13 other individual databases were mentioned, including 3 proprietary repositories maintained by the applicants.

Notably, more than one-third (9) of the plans that included data deposit provisions did not commit to use a particular database. The most general statement was that data would be deposited in “standard and freely available databases.” Another plan pledged to deposit data in an “appropriate” repository to be selected by the project steering committee. The most common approach, taken by six of the plans, was to reference one or more of the repositories named in the PADS Policy guidance example as options, but without committing to any specific repository. Finally, one plan indicated that the investigators intended to develop a new data sharing platform as part of the proposed project. This plan provided some platform specifications including user registration and authentication, but most of the supporting details discussed data formatting.

The Cancer Moonshot program anticipated that data would be shared via a federally managed National Cancer Data Ecosystem (Cancer Moonshot Blue Ribbon Panel Report, 2016). To this end, some Cancer Moonshot funding announcements indicated that data sharing and coordination would be required with NIH data repositories including the Cancer Epidemiology Data Repository (CEDR) (Department of Health & Human Services (HHS), 2017). Yet, as of this writing, the CEDR data repository has still not been launched, leaving some grantees unable to satisfy the commitments made in their PADS plans.

Responding to the OMB directive of 2013, the NCI PADS Policy also asks applicants to specify how their plans would address “the methods through which the public including other researchers will locate and access the data.” Among the 17 plans that addressed this element, the most common offerings were direct sharing when data is requested by another researcher (9 plans), and a commitment in 8 plans to make presentations at conferences and other scientific gatherings. In addition, 7 plans stated that researchers would publicize information about their Cancer Moonshot research on a public website, with 6 of these plans making reference to the website including information about how to access or request data.

Discussion

Overall assessment

The above findings suggest that, during the first year of its effectiveness, there were significant gaps in the implementation of the Cancer Moonshot PADS Policy. Though the submission of a PADS plan addressing both publication and data sharing was required in all applications for Cancer Moonshot funding, nearly half of the 33 PADS plans from funded grant applications that we reviewed did not engage meaningfully with both prongs of the PADS Policy, and 8 did not mention any of the PADS Policy elements for publications or data sharing.

Sixteen plans did not mention open access publication. Of those that did, several were cursory and nonspecific. In a similar vein, though the PADS Policy asked applicants to specify how their plans would address “the methods through which the public including other researchers will locate and access the data,” nearly half did not, and most of those that mentioned this policy element did so without making definitive, detailed commitments.

In terms of data repositories, applicants mentioned a total of 16 different databases in which data would be deposited, including 3 proprietary repositories. We are aware of no plan by NCI or others to enable the interoperability or long-term stewardship of these data repositories. Such a diverse array of independent

data repositories is not likely to facilitate the efficient public utilization of data generated under the Cancer Moonshot Program (Contreras and Reichman, 2015; Sherkow, 2018).

As discussed above, one-third of the plans submitted to NCI referenced at least one earlier NIH data sharing and publication policies from 1999, 2003, and 2008, even when the PADS plan superseded the requirements of these earlier policies. This suggests that these plans were, at least to some degree, copied from earlier grant applications that pre-dated the PADS Policy. In fact, a majority of the plan language that we reviewed appeared to originate from grant applications that were not subject to the PADS Policy.

Analysis

The degree to which successful Cancer Moonshot grantees—many of which were experienced research teams from leading institutions—omitted reference to the PADS Policy suggests that grant applicants were uncertain about the requirements of the PADS Policy, or did not appreciate that the PADS Policy would apply to their research. We suspect that some terms in the PADS Policy, such as “CC BY 4.0” and “public domain,” which may be familiar to those in the policy community, may be less familiar to researchers accustomed to publishing their work in traditional scholarly journals and to grant managers and science writers who may have been asked to assist with the preparation of this portion of the Cancer Moonshot applications. Journals themselves may be unfamiliar with the aggressive publication sharing principles of the Cancer Moonshot Program (Kaiser, 2019b). While *Cell* and *Science* journals have now adopted special open access provisions for Cancer Moonshot-funded articles, others, such as *Nature*, will consider requests on a case by case basis only (Kaiser, 2019b). And beyond these few top-tier journals, the vast majority of medical and scientific journals have no provision for rapid release of Moonshot-funded articles.

Likewise, though the precepts of scientific data sharing have been acculturated among researchers in genomics and related fields, they may be less familiar to researchers seeking Cancer Moonshot funding in fields such as oncology, surgery, pharmacology, and social sciences. Moreover, as demonstrated by the unavailability of repositories for nongenomic data types, such as clinical, social, and behavioral data, it appears that NCI did not adequately prepare for the rollout of such an ambitious data sharing program across multiple scientific disciplines.

Given the significant changes to existing data sharing practices that were introduced by the PADS Policy, the prominence with which the policy was announced by NCI, and the likely inexperience of many potential Cancer Moonshot grant applicants with the intricacies of publication and data sharing policy (including, e.g., Creative Commons licensing), significant NCI planning, training, and educational outreach to the applicant community might have been expected with regard to the PADS Policy. Yet little was offered in this regard. Neither the PADS Policy nor NCI’s briefing materials explain how the requirements and recommendations of the PADS Policy differ from prior NIH publication or data sharing policies, requiring researchers accustomed to working within existing frameworks to assess policy differences on their own. In the NCI-sponsored webinar announcing and explaining one of the FY18 Cancer Moonshot grant opportunities, of 17 slides that were presented, only one-half of one slide included information about publication and data sharing, and this simply referenced an existing NCI genomics repository (GDC) and the still pending epidemiology data repository mentioned above (CDER) (National Cancer Institute (NCI), 2017b).⁵ Notably, the slides describing the four “Goals” and five “Aims” of the Cancer Moonshot program did not mention data or publication sharing, or the PADS Policy, at all (National Cancer Institute (NCI), 2017b). Likewise, neither publication/data sharing nor the PADS Policy

⁵ Slide 10 contains the following text:

1. Required data/resource sharing with existing NIH resources.
 - 1.1. Submit a Pubic [sic] Access and Data Sharing Plan.
 - 1.1.1. Genomic Data Commons (GDC).
 - 1.1.2. Cancer Epidemiology Data Repository (CDER).

are mentioned in the NCI FAQ relating to this funding mechanism (National Cancer Institute (NCI), 2017c).

As a result, we suspect that NCI's aspirational data sharing initiatives, conceived at the highest levels of the Institute, may not have been adequately disseminated within NCI. In particular, we believe that more practical education regarding the PADS Policy, and its publication/data sharing mechanisms and expectations, should have been offered to NCI program officers overseeing Cancer Moonshot funding mechanisms, as these individuals are often the first line of inquiry when investigators are faced with complex new application requirements. In addition, it is possible that NCI's Cancer Moonshot study sections, comprised of volunteer researchers from across the country, were also unfamiliar with the intricacies of the PADS Policy. In order for these individuals, as well as the members of NCI's Council, which ultimately approves all grant funding, to assess grant applications in a manner that is consistent with NCI policy, additional training and guidelines for the review of applications subject to the PADS Policy would have been helpful for this new area of "funding priority."

Broader implications

The PADS Policy, in an effort to offer flexibility to researchers dealing with a broad spectrum of data types, permits grant applicants to submit a PADS plan for review and assessment by NCI. However, as shown above, this degree of flexibility was not effective when applicants were inadequately apprised of the importance or details of the PADS Policy, and when NCI itself had not adequately considered the mechanisms for sharing of data types beyond traditional genomic and associated phenotypic data.

Similar to the PADS Policy, NIH's new draft Policy for DMS also relies on applicant-generated data sharing plans rather than concrete data sharing requirements. This reliance on applicant-generated plans, rather than concrete agency mandates, could result in plans that overlook or undervalue data sharing, a concern that we share with other commenters on NIH's draft DMS Policy (Sim et al., 2020). By the same token, without concrete guidance from NIH, it is not reasonable to expect individual research teams to formulate and deploy adequate data sharing plans for novel data types. Moreover, this individualized plan approach, by pushing the work of developing data sharing practices for diverse data types onto investigators rather than funding agencies, eliminates any prospect of uniform and interoperable data resources that are usable broadly by the research community. If data sharing migrates from the centralized repositories currently operated by NIH (through the National Center for Biotechnology Information) in the genomics arena to a multitude of fragmented institutional repositories, the overall benefit to the scientific community is likely to diminish (Contreras and Reichman, 2015; Sherkow, 2018).

Our findings suggest a need for more robust planning, communication, and education regarding publication and data-sharing requirements. Offering concrete guidance and expectations regarding the logistics and practicalities of sharing both genomic and nongenomic data is critical. The need for training is urgent, both with respect to potential grant applicants and also the many program officers throughout NIH who will be expected to explain and manage these new policies. These points will become particularly important as data sharing requirements move from the relatively circumscribed realm of genomics and related fields to the vast array of health science disciplines covered by NIH-funded research with the eventual goal of forming an integrated "medical information commons" (Bollinger et al., 2019). For researchers in these disciplines, the aggressive data sharing precepts that originated with the HGP may be both unfamiliar and counterintuitive, and significant acculturation may be required in order to ensure the successful implementation of these data sharing practices across the scientific community.

Conclusion

NCI's Cancer Moonshot PADS Policy was adopted with the ambitious goal of "accelerat[ing] cancer research by enhancing data sharing and making evidence-based approaches to cancer therapy more broadly available to patients." To this end, NCI's PADS Policy required the submission of a PADS plan

with each funding application under the Cancer Moonshot Program, and emphasized the “funding priority” would be given to applicants that ensured “maximal sharing” of resulting publications and data.

Yet our review of 33 PADS plans submitted with funded FY18 Cancer Moonshot applications reveals that few fully addressed the requirements of the PADS Policy. Many, in fact, appeared to rely on pre-PADS language taken from prior grant applications that referenced NIH policies dating from 1999 to 2003. Moreover, there is little evidence that NCI evaluated grant applications based on their PADS plans as submitted without subsequent guidance from NCI program staff, given how many were funded after submitting partial information about publication and data sharing plans.

We believe that these gaps arose from a combination of insufficient planning, education, and outreach by NCI concerning its PADS Policy, both with respect to potential grant applicants and NCI’s program staff and external grant reviewers. Accordingly, we recommend that NCI undertake improved planning and educational efforts regarding the importance and implementation of the PADS Policy for relevant Cancer Moonshot program officers and study section members. And once the applicant community is adequately apprised of the data and publication sharing requirements of the PADS Policy, we hope that NCI diligently enforces its “funding priority” to select applications that ensure “maximal sharing” of publications and data.

We believe that our findings and recommendations are relevant beyond NCI and the Cancer Moonshot, as NIH is currently considering the adoption of a similar plan-based data sharing policy that would apply across all NIH institutes and centers. Accordingly, we would urge NIH to consider the recommendations made above with respect to its draft DMS Policy.

Acknowledgments. The authors acknowledge data coding assistance by Tim Nielsen, as well as administrative support from the Center for Law and Biomedical Sciences (LABS) at the University of Utah S.J. Quinney College of Law. Helpful suggestions on this article were provided by Prof. Jacob Sherkow and two anonymous reviewers.

Funding Statement. This project was funded in part through a pilot grant awarded by the Utah Genome Project (effective March 1, 2018). J.L.C. has also received support from the Huntsman Cancer Institute at the University of Utah. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests. The authors declare no competing interests exist.

Authorship Contributions. Conceptualization, T.M.F. and J.L.C.; Methodology, T.M.F. and J.L.C.; Formal Analysis, T.M.F. and J.L.C.; Data Curation, T.M.F.; Writing-original draft, T.M.F. and J.L.C.; Writing-review & editing, T.M.F. and J.L.C.; Supervision, J.L.C.; Funding Acquisition, J.L.C.

Data Availability Statement. The data that supports the findings of this study were coded by the authors following our original collection of NCI Cancer Moonshot applications and data sharing plans directly from a sample of Cancer Moonshot grantees. We promised grantees anonymity for their participation in this study. Because it is not possible to de-identify the submitted plans or the coded portions, given that plan content often refers to details of the proposed research project, we are unable to provide plan text beyond the descriptive statistics included in this article.

References

- Bollinger JM, Peter D. Zuk, Mary A. Majumder, Erika Versalovic, Angela G. Villanueva, Rebecca L. Hsu, Amy L. McGuire, and Robert Cook-Deegan** (2019) What is a Medical Information Commons? *Journal of Law, Medicine & Ethics* 47:41–50.
- Cancer Moonshot Blue Ribbon Panel Report** (2016).
- Contreras JL** (2011) Bermuda’s legacy: Patents, policy and the design of the genome commons. *Minnesota Journal of Law, Science, & Technology* 12(1), 61–125.
- Contreras JL and Knoppers BM** (2018) The genomic commons. *Annual Review of Genomics & Human Genetics* 19, 429–453.
- Contreras JL, Reichman JH** (2015) Sharing by design: Data and decentralized commons. *Science* 350(6266), 1312–1314.
- Creative Commons** (2019) Creative Commons Licenses. Available at <https://creativecommons.org/use-remix/cc-licenses/> (accessed 16 August 2019).
- Department of Health & Human Services (HHS)** (2017) RFA-CA-17-041: Approaches to Identify and Care for Individuals with Inherited Cancer Syndromes (U01).
- Fleming K** 2019 Email Announcement from Karen Fleming (NIH/NCI) “Moonshot Open Access Policy” (4 November 2019).

- International Human Genome Sequencing Organization** (1996) Summary of Principles Agreed Upon at the First International Strategy Meeting on Human Genome Sequencing (Bermuda, 25–28 February 1996) as Reported by HUGO. Available at http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml accessed date June 13, 2020.
- Kaiser J** (2019a) Open access takes root at National Cancer Institute. *Science* 365(6454), 629.
- Kaiser J** (2019b) In departure for NIH, Cancer Moonshot requires grantees to make papers immediately free. *Science News* <https://doi.org/10.1126/science.aaz1344>.
- Mattioli M** (2017) The data-pooling problem. *Berkeley Technology Law Journal* 32, 179–236.
- National Cancer Institute (NCI)** (2017a) NCI Cancer Moonshot Public Access and Data Sharing Policy, 4 August 2017. Available at <https://www.cancer.gov/research/key-initiatives/oonshot-cancer-initiative/funding/public-access-policy> (accessed 18 November 2018).
- National Cancer Institute (NCI)** (2017b) Pre-Application Funding Opportunity Announcement (FOA) Webinar: Approaches to Identify and Care for Individuals with Inherited Cancer Syndromes, 4 October 2017. https://healthcaredelivery.cancer.gov/media/Hereditary_cancer_RFA_webinar.pdf
- National Cancer Institute (NCI)** (2017c) Pre-Application Webinar on Approaches to Identify and Care for Individuals with Inherited Cancer Syndromes (RFA-17-041) Frequently Asked Questions https://healthcaredelivery.cancer.gov/media/Inherited_Cancer_Syndromes_FAQ_FINAL.pdf.
- National Cancer Institute (NCI)** (2019) NCI Cancer Moonshot Research Initiatives. Available at <https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/implementation> (accessed 3 December 2019).
- National Institutes of Health (NIH)** (1999) NIH Research Tools Policy (principles and guidelines for recipients of NIH research grants and contracts on obtaining and disseminating biomedical research resources). *Federal Registry* 64(246), 72090–72096.
- National Institutes of Health (NIH)** (2003) Final NIH Statement on Sharing Research Data, NOT-OD-03-032, 26 February 2003. Available at <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.
- National Institutes of Health (NIH)** (2008) Public Access Policy implementing Division G, Title II, Section 218 of PL 110-161 (Consolidated Appropriations Act, 2008).
- National Institutes of Health (NIH)** (2014) Final NIH Genomic Data Sharing Policy. *Federal Register* 79(167), 51345–51354.
- National Institutes of Health (NIH)** (2015) Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research. Available at <https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>.
- National Institutes of Health (NIH)** (2019) Request for public comments on a DRAFT NIH policy for data management and sharing and supplemental DRAFT guidance. *Federal Register* 84(217), 60398–60402.
- Office of Management and Budget (OMB)** (2013) Memorandum for the Heads of Executive Departments and Agencies: Open Data Policy—Managing Information as an Asset (M-13-13).
- Sherkow JS** (2018) Cancer’s IP. *North Carolina Law Review* 96, 297–380.
- Ida Sim, Michael Stebbins, Barbara E. Bierer, Atul J. Butte, Jeffrey Drazen, Victor Dzau, Adrian F. Hernandez, Harlan M. Krumholz, Bernard Lo, Bernard Munos, Eric Perakslis, Frank Rockhold, Joseph S. Ross, Sharon F. Terry, Keith R. Yamamoto, Deborah A. Zarin, Rebecca Li** (2020) Time for NIH to lead on data haring. *Science* 367(6484), 1308–1309.