AcademyHealth Comments on the NIH Draft Data Management and Sharing Policy and Supplemental Guidance

AcademyHealth consulted with a committee of members and thought leaders to offer a response to NIH's request for comments on their draft Policy for Data Management and Sharing. We thank and acknowledge Greg Downing for his valuable contributions and guidance as Chair of the Ad-Hoc Committee.

AcademyHealth represents 4,000 individuals and organizations in the research community using evidence and data to improve health and health care for all. Our organization recognizes the crucial role of data sharing in advancing scientific research, and ultimately improving clinical care and patient outcomes. We commend the National Institutes of Health for their efforts to optimize access to and use of shared data in research. Below, we provide our comments and suggestions on NIH's draft Policy for Data Management and Sharing. These comments intend to build on and reinforce our feedback provided in December 2018 on NIH's proposed provisions for this policy. At a high level, our feedback offers suggestions for enhancing clarity and specificity with respect to terminology, requirements, and components of a data management and sharing plan (Plan), and perspective on the use of resolute language to effectively communicate NIH's expectations, and commitment to upholding Plans to rigorous scientific standards.

Section I: Purpose
AcademyHealth believes it is crucial to pay careful attention to context and rationale provided for the purpose of data sharing. The argument to support enhanced data sharing must be compelling and it needs to resonate with constituents to garner support. The first sentence of the policy describes NIH's commitment to "making results and outputs of the research it funds and conducts available to the public." We would argue that the scope of data management and sharing should be broadened to encompass data creation and collection, and not solely focus on results and research outputs. At the outset, the policy should state not only the role of shared data in optimizing research results, but also underscore broader implications for the field, including enhanced collaboration, transparency and accountability.

In terms of setting context, how the benefits of data sharing are communicated, including the order in which they are described, are important. Currently, "to test the validity of research findings" is listed as the first (and presumably most important) benefit of data sharing. We would argue the benefits may be better expressed tied to the goals of NIH, and the value of reusing data for deeper or broader discovery should be the emphasis.
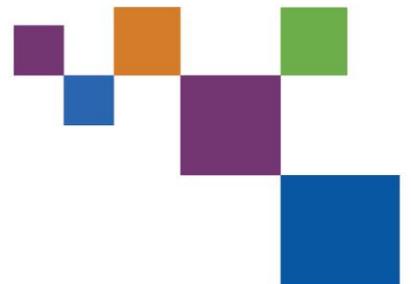
Further, while FAIR principles are valuable for considering data sharing, it is unclear whether investigators subject to the proposed policy for data management and sharing are responsible for each of these principles. We believe that the focus of sharing should be to make the data accessible and interoperable, and the role of the NIH is to make data findable and usable. As currently written, investigators sharing the data take on the full burden, which may be counter-productive. We suggest revising the plan such that it diminishes investigator burden through technical, infrastructure and cost sharing across the spectrum of stakeholders in the publication and sharing of data.

Section II: Definitions
Data Management and Sharing Plan (Plan): The definition of the Plan in the proposed draft should be considered more broadly. Rather than just defining how the data will be "managed, preserved and shared," it is important that the Plan also identify at least one target "other" who would use the data, and describe how that user would actually use it. Otherwise, the Plan will only describe how to make it possible, rather than to enable it.

The definition also does not address how the scientific data will be collected or described in the shared data set, which are important components of the Plan.

Data Management: No comment
Data Sharing: The proposed definition for data sharing is limited, and should more explicitly state that sharing is more than increasing access. Sharing data must involve enabling the access and reuse of data to facilitate and optimize research. In addition, because it is important to understand how results were determined, and considering that current data analytics provide unique methods of data analysis and interpretation, sharing the associated code that was used to determine the accuracy of analyses should be required.

Metadata: No comment
Scientific Data: The proposed definition of scientific data explicitly excludes "preliminary analyses" as eligible material. However, many data will be rendered futile for the purposes defined in data sharing without a clear understanding of preliminary analyses, and in some cases access to that data. At a minimum, the definition could state the scientific data "may exclude" preliminary analyses, rather than an unequivocal exclusion.

Section III: Scope
We believe an important opportunity inherent in this policy is to expand access to data extracted from electronic heath records for secondary analyses, which represents a growing area of important research. It is possible that many studies using secondary data analysis will not explicitly de-identify the data if not legally required (e.g., if the data do not leave the institution). To mitigate this potential limitation, the scope should be broadened to recommend the creation of shareable data sets to support these analyses.

Section IV: Effective Date(s)
No comments.

Section V: Requirements
NIH guidance on the submission of the Plan should be as specific and prescriptive as possible. For example, the "Requirements" section notes that the submission of the Plan should outline how scientific data will be managed and shared, "taking into account any potential restrictions or limitations." Instead of taking into account potential restrictions or limitations, the Plan should clearly describe what those restrictions and limitations are. For example, if the data contains proprietary information that imposes restrictions on sharing, this should be clearly described in the Plan prior to NIH funding decisions. The "Requirements" section should clearly state which elements of the Plan are required and which are optional.

We agree with the statement that "additional or specific information" may be requested to meet expectations for data management and sharing. NIH may consider elaborating on this statement and stating that the creation of shareable de-identified data sets may be requested (even if not required for observational studies). We encourage NIH to consider how to ensure sufficient resources (budget) exist to support that aspect of the Plan.

Section VI: Data Management and Sharing Plans
AcademyHealth members recommend several suggestions related to the Plan, which are organized around topic area below.

> *Data security and privacy.* The draft policy suggests that investigators are responsible for ensuring data security and compliance with privacy protections throughout the life of the scientific data, even after it has been shared. This is a tremendous and daunting requirement. Further, with changing discoveries in security and privacy protections, this requirement may be beyond the capability of most researchers and institutions, and may limit sharing. One strategy to alleviate this burden would be the use of NIH repositories (see next subsection for additional thoughts and comments).

> *Use of repositories*. The draft policy states that NIH "encourages the use of established repositories for preserving and sharing scientific data." NIH should elaborate on the utility of repositories, why repositories are encouraged, and the criteria for establishing a repository. Given the statement that Plans should identify strategies or approaches to ensure data security and compliance with privacy protections through the life of the data, repository use could be incentivized (e.g. by making administrative supplements available, providing the investigator extra "points" for a track record of using repositories as part of subsequent grant reviews, etc…), not just encouraged. Use of established repositories or repositories within NIH would lift the burden of privacy and security protections from the researcher and be assumed by NIH, who can better represent the public need for the data.

Plan Elements. The proposed guidance suggests that investigators "consider" addressing specific Plan elements outlined in the supplemental guidance. We believe that requiring applicants to address all elements listed in the supplemental guidance would provide clarity to applicants on expectations for an adequate Plan as well as assist NIH with their review of the Plan.

Making plans public. The policy states that NIH may make Plans publicly available. Revising this statement to indicate that NIH *will* publish Plans for public consumption removes any uncertainty that this may or may not happen, and promotes transparency and accountability among the community of NIH researchers.

Peer review. As currently written, the proposed policy does not require that Plans be evaluated as part of the peer review process. This separation from Peer Review suggests that specific elements of the Plans are not subject to an acceptable level of scientific scrutiny, and therefore may not carry as much importance as the rest of the study plan. We believe that the Plan for making data accessible for evaluation and further research should be regarded with the same level of importance and integrity as the study plan, and note that doing so also requires the engagement of reviewers with appropriate knowledge and expertise to evaluate the Plan.

Section VII: Compliance and Enforcement
No comments.
Supplemental DRAFT Guidance: [Allowable Costs for Data Management and Sharing](#)
NIH's guidance on allowable costs for data management and sharing should include generating shareable datasets when legal privacy protections would otherwise restrict the sharing of data. NIH should also collect information about the anticipated costs to researchers to access the study's scientific data in the intended repository. Data that will carry higher than average access costs become essentially inaccessible.

Supplemental DRAFT Guidance: [Elements of a NIH Data Management and Sharing Plan](#)
AcademyHealth members recommend additional suggestions related to elements of the Plan, which are organized around topic area below.

Introductory Language. The draft guidance states that NIH does not expect researchers to share all scientific data generated in a study. NIH should provide more clarity on what types of specific scientific data researchers are not expected to share. AcademyHealth feels strongly that scientific data created through NIH funding that will not be shared should include a justification for its exclusion.

Data Type. Descriptions of the modality, level of aggregation, and degree of data processing should be considered basic requirements of the Plan. These descriptions would form the minimally necessary metadata that will make the scientific data understandable to others. Without this information, other researchers would be burdened with producing this basic level of information, impeding and delaying further use and reuse, slowing the progress of science, and undermining the objective of data sharing.

Access to Data. With regard to access to restricted scientific data, the applicant should be required to explain their process for obtaining approval to the restricted data. NIH should be able to determine from the Plan how likely it is that others will also be able to access the restricted data.

Other Use Limitations. There should be specific consideration of limitations imposed by sharing protected health information and identifiable information. These are significant limitations and should be addressed directly, along with proposed strategies to mitigate these challenges.

Other Considerations Relevant to this DRAFT Policy Proposal
The draft policy alludes to the significance of the HIPAA Privacy Rule, but does not explicitly address the implications for data management and sharing. The HIPAA Privacy Rule identifies research as a public interest and benefit activity, and if such allowances are made, the Plan should include how to address protected health information. Otherwise, it may be easy to exclude research using protected heath information from this policy.

An additional element and benefit of data sharing is the importance of open data to spark cross-disciplinary research. Members noted that the availability of valid data resources is a key strategy for bringing new investigators from other fields into health and health care research. We believe the guidance should be prescriptive enough to provide funding panels clear insight into the likely success of the research project in terms of data sharing-- that is, the likelihood that the study's data could be used to validate or replicate study findings or be reused for further research. The NIH should make it clear to researchers that insufficient data management and sharing plans will affect the likelihood of receiving funding.

Finally, beyond and in addition to the whole of our comments above, we want to reemphasize the importance of providing centralized support – infrastructure, operational guidance, resources, logistical support and training – to the field in order to support meaningful implementation of this policy.

Conclusion
AcademyHealth appreciates the opportunity to provide additional input on the Draft Policy for Data Management and Sharing. We believe clarity and specificity will be paramount to a sound and effective policy – one that clearly communicates context, rationale, requirements, and expectations and advances progress toward the widespread and responsible sharing of data.