December 10, 2018

Carrie D. Wolinetz, Ph.D.
Associate Director for Science Policy
NIH Office of Science Policy
National Institutes of Health
6705 Rockledge Drive, Suite 750
Bethesda, MD 20892

Re: Proposed Provisions for a Draft NIH Data Management and Sharing Policy

Dr. Wolinetz:

AMIA's membership is comprised of informaticians across the spectrum of biomedical research, clinical care, public health, and consumer health, with backgrounds in medicine, biomedical sciences, and informatics. Our comments are rooted in this expertise and are representative of diverse and multidisciplinary stakeholders who are deeply experienced in the systematic collection, analysis, application and responsible sharing of data for health.

AMIA enthusiastically supports development of a pan-NIH Data Management and Sharing Policy (DMSP) and we commend the NIH for initiating this effort. We are pleased to see several elements of AMIA's Data Sharing Principles & Positions incorporated in the Proposed Provisions, including a reliance on FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles, and acknowledgment that the DMSP should support underlying infrastructure and curation activities with funding.

We are especially pleased that the NIH envisions a DMSP that applies to "all intramural and extramural research, funded or supported in whole or in part by NIH, that results in scientific data, regardless of NIH funding level or mechanism." While this scope is ambitious, quality data management and sharing plans (Plans) are prerequisite to achieve the vision of FAIR data principles and such a scope should be the long-term goal of the NIH DMSP.

Recognizing the need to have all NIH-funded research comply with this DMSP, and with appreciation of what AMIA sees as necessary components of a data management and sharing plan, we recommend a phased compliance timeline based on funding levels. This phased implementation would only apply to new research funded after the DMSP is final. First, new research funded above $500,000 per year and subject to the existing data sharing policy should comply with the final DMSP within one year of its adoption. Second, new research funded above $250,000 per year should comply with the provisions of the DMSP within 2 years of its adoption, and finally, all grants funded below $250,000 per year should comply with the DMSP within 3 years of adoption. This compliance approach would focus efforts on those grants that already must comply with the existing policy and

likely have the richest cache of scientific data, while giving smaller projects more time to become familiar with the DMSP.

Alongside this phased adoption timeline, the NIH should consider a graduated DMSP that appropriately calibrates requirements based on funding level and whether scientific data are deposited in an NIH-endorsed depository or knowledgebase. We strongly recommend that the draft DMSP encourage Institutes and Centers (ICs) to factor the quality of the Plan into the overall impact score through the peer review process for those grants that are supported at high levels or support programmatic priorities. We also recommend that NIH incentivize deposition of scientific data in NIH-endorsed databases and knowledgebases by allowing such Plans to comply with a streamlined DMSP.

We note several high-level observations and recommendations for which we provide additional detail and rationale in the enclosure of this comment letter:

1. **The draft DMSP should improve data management and sharing of scientific data to facilitate learning health systems and continuous discovery.**
    a. While we are supportive of a pan-NIH DMSP, subject to ICs specific grant-types and awardees, AMIA recommends the DMSP encourage ICs to make Plans scorable elements of specific grants. This will improve Plans' quality and better ensure supplemental use of scientific data.
    b. AMIA also recommends the DMSP seeks to improve the interoperability and supplemental uses of research data writ large by encouraging the use of established biomedical data standards and adherence to data management and data sharing best practices. Over time, better use of and refinement of data standards, buttressed by systematic scoring of plans, will optimize scientific data for continuous learning and discovery.
    c. AMIA recommends the DMSP incentivize the deposition of scientific data and tools, software and/or code developed as part of NIH-supported projects into NIH-approved data repositories and knowledgebases. This will enable both large and small grantees to more easily comply with the DMSP.
2. **The draft DMSP should improve institutional support and professional advancement for experts managing and sharing scientific data.**
    a. We applaud NIH for suggesting that reasonable costs associated with data management and sharing could be requested under the budget for the proposed project. AMIA recommends that the DMSP establish a standard way to account for data management and sharing costs as both Direct costs and F&A costs.
    b. The DMSP should facilitate implementation of the NIH Data Science Strategic Plan, especially the relevant aspects of the Strategic Plan that seek to credit experts who manage and share valuable data sets / software for their work. If data is seen as valuable, experts who enable FAIR data should also be valued. The NIH should support certifications for experts that manage and share scientific data. We also see a need for R&D on data management tools to facilitate compliance with the DMSP.
3. **To operationalize the DMSP more specificity and clarity around concepts is needed.**
    a. Data management is distinct from data sharing. The processes and activities that support data management and sharing are also different. AMIA recommends the

NIH develop a DMSP that specifies these distinctions through additional Plan Elements as described below.

b. AMIA recommends that the DMSP expand the current list of definitions to include concepts for "Data Management," "Covered Data," "Covered Timeframe," and refine definitions for "Metadata" and "Scientific Data."
c. While we support the scope of a pan-NIH DMSP that covers all grants, contracts, and/or other funding agreements, AMIA recommends the NIH convene stakeholders with individual ICs to operationalize the DMSP.

Finally, we offer AMIA and its members as resources during subsequent work on the DMSP. We strongly recommend the NIH develop a subsequent draft DMSP based on stakeholder feedback to the concepts in this RFI. Another comment period will provide NIH with valuable insights before issuing a final DMSP.

The enclosure includes detailed AMIA comments regarding the Proposed Provisions for a Draft NIH Data Management and Sharing Policy. Where possible, we provide both in-line edits and rationale for suggested edits.

We hope our comments are helpful as you undertake this important work. Should you have questions about these comments or require additional information, please contact Jeffery Smith, Vice President of Public Policy at jsmith@amia.org or (301) 657-1291. We look forward to continued partnership and dialogue.

December 10, 2018

Sincerely,

Douglas B. Fridsma, MD, PhD, FACP, FACMI
President and CEO
AMIA

Enclosed: Detailed AMIA comments regarding the Proposed Provisions for a Draft NIH Data Management and Sharing Policy.

December 10, 2018

I. Definitions

a. Data Management and Sharing Plan

AMIA Comments: The draft Data Management and Sharing Policy (DMSP) should differentiate between "data management" and "data sharing" as two distinct concepts and sets of activities with different, if overlapping, considerations and timeframes. For clarity, we refer to the Data Management and Sharing Plan as "Plan" and DMSP refers to the policy. While we are supportive of the focus on data sharing as part of data management, it is critical to acknowledge that upstream data collection and handling processes largely determine data quality necessary for research replicability, reproducibility, and traceability.[1]

AMIA Comments: The draft DMSP should not distinguish between potential "others" who may use scientific data. The number and heterogeneity of "others," even when confined to "researchers," and "the broader public," would needlessly complicate compliance with the DMSP. AMIA members note a major discrepancy between making scientific data available to another scientist in the same discipline and making it available to the general public. Further, we note that even within the scientific community, there will be wide gaps in knowledge across disciplines that requires extensive annotation and training to be understood.

AMIA Recommendation: AMIA recommends the following amendments to the Plan's definitions to acknowledge differences in data management and data sharing. Further AMIA recommends the draft DMSP remove all references to "(e.g. researchers and the broader public)" when describing potential users of scientific data:

Data Management and Sharing Plan: A plan describing how scientific data will be generated, managed, described, analyzed, preserved, shared, and made accessible to others for supplemental uses, ~~(e.g., other researchers and the broader public),~~ as appropriate. This plan should include two distinct sections describing how scientific data will be managed across the life-cycle of the project and how scientific data will be shared at the project close, or at another appropriate interval(s).

b. Data Management

AMIA Comments: As discussed above, the DMSP should explicitly describe what is necessary to manage data, not just share data, given that data management and data sharing are distinct. Data management is prerequisite for data sharing, ensuring that the data are accurate, complete, and maintained in a standardized manner. Without effective data management, you cannot have effective data sharing, thus we recommend the DMSP consider additional Plan Elements as described in that section of our comments.

AMIA Recommendation: Given this view, we recommend the draft DMSP include a new definition for data management as follows:

[1] Traceability of research data is the ability to reproduce the raw data from the analysis datasets and vice versa.

December 10, 2018

Data Management: The upstream management of scientific data that documents actions taken in making research observations, collecting research data, describing data (including relationships between datasets), processing data into intermediate forms as necessary for analysis, integrating distinct datasets, and creating metadata descriptions. Specifically, those actions that would likely have impact on the quality of data analyzed, published, or shared.[2]

c. Data Sharing

Data Sharing: ~~To make~~ Making scientific data accessible for use by others ~~(e.g., other researchers and the broader public)~~ in a manner that is consistent with the FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles.

d. Metadata

AMIA Comments: We found the definition for metadata in need of refinement. Specifically, the phrase "additional information to make data more usable" implies that a data set could be usable at all without metadata, which is simply not the case. There is no data that can be correctly understood, much less re-used, without at least a data dictionary with field definitions and data types. Further, we view "Outcome measures" as actual data, not metadata. There may be metadata that defines how an outcome measure was derived, but the outcome data itself is not metadata.

AMIA Recommendation: Given this view, we recommend the draft DMSP amend the definition of metadata as follows:

Metadata: ~~Data that provide additional information to make data more usable (e.g., independent sample and variable description, outcome measures, and any intermediate, descriptive, or phenotypic observational variables).~~ Metadata is descriptive information about data, including variable/document definition/description, data type, and other characteristics. Areas discussed in metadata include, but are not limited to, instruments used to collect data; parameters or settings for such instruments; descriptors of physical samples from which data were collected; dates and times of data collection; any transformations applied to the data; relationships between datasets; provenance linking derived or modified datasets to original sources; phenotypic descriptors of data sources; and institutional/personal identifying information associated with the group or person(s) responsible for the data. Metadata also help establish (confidence in) the credibility of the data.

In survey data, "paradata" is used to describe confidence in the credibility of data. This may be an evaluation of the sincerity or seriousness of the respondent by the questioner (e.g. "Open/Frank" to "Uncomfortable/Evasive", "Earnest" to "Flippant", etc.), or less subjectively, in an online survey, the time the respondent spent to complete the survey.

[2] Adapted from Williams, Bagwell and Zozus "Data management plans, the missing perspective," Journal of Biomedical Informatics 71 (2017) 130–142

e. Scientific Data

AMIA Comments: We support the concept of "Scientific Data," but do not support a definition of this concept through negation. The listing of what Scientific Data is not may serve better as part of ancillary materials published by the NIH, such as Frequently Asked Questions, rather than be included in a definition. Further, it is odd to place a command, "NIH expects…" into a definition.

AMIA Recommendation: Given this view, we recommend the draft DMSP include a new definition for Scientific Data as follows:

Scientific Data: ~~The recorded factual material commonly accepted in the scientific community as necessary to validate and replicate research findings including, but not limited to, data used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens. For the purposes of a possible Policy, scientific data may include certain individual level and summary or aggregate data, as well as metadata. NIH expects that reasonable efforts should be made to digitize all scientific data. 1.~~ Information that is gathered, derived or generated in the course of conducting research. It is the basis for reaching conclusions and inferences based on scientific principles and methodologies. Scientific data can be used to test existing hypotheses, to generate new hypotheses for future research, to validate or replicate prior research as well as for more exploratory purposes. Scientific data represent the foundation for both scientific theories and publications.

f. AMIA Recommended New Definitions

1. Scientific Software Artifacts

AMIA Comment: Increasingly, the NIH funds research that results in software tools, code, and analytic programs. These "software artifacts" are both explicitly funded as part of extramural research and developed as a means to conduct NIH-funded research. These software artifacts can be deposited in knowledgebases analogous to data into databases.

AMIA Recommendation: AMIA recommends the draft DMSP includes a definition for "Scientific Software Artifacts," so that grantees clearly understand that both data and software tools created with NIH funds should be included as part of their data management and sharing plan. This definition would be limited to artifacts created with NIH funds, and omit proprietary software tools used to conduct research, such as a stat package. We recommend a definition such as:

Scientific Software Artifacts: Software, code, analytic programs, and other knowledge artifacts developed to conduct research or resulting from the conduct of research.

2. Covered Data

AMIA Comment: We see the need to define two additional terms so that the DMSP can address a number of questions that arise throughout our deliberations. Specifically, grantees need to have a

clear understanding of which Scientific Data are covered by the Policy and for what period of time those data are covered. These definitions do not need to establish a policy for these questions; rather, these concepts should facilitate conversations to answer those questions.

There is a distinction between data generated by and for research, and data that is used in research. We see a need to define what scientific data is covered under the DMSP and what data is not. For example, clinical trials routinely rely on data that has been generated during the course of clinical care and collected as part of research participants' electronic health record (EHRs). This data may be included in the study data set and used as part of an analysis. Such data was not specifically generated for the trial and the tests or other work involved in generating them were not paid for by the trial. Is such data covered by the policy or not?

As another example, we note that a number of large databases are currently used and made available for epidemiological research or data mining projects based entirely on real-world evidence. These data are generated and paid for in the course of routine clinical care and are maintained under private funding. If an NIH funded analytic project is based on the use of such data, can that data now be required to be made available more generally to the public? If so, this could represent a disincentive to the private organization to make such data available for research and might have the paradoxical effect of making less data available for research or making whole classes of data unavailable for research.

AMIA Recommendation: AMIA recommends the draft DMSP includes a definition for "Covered Data," so that grantees clearly understand which data must be included as part of their data management and sharing plan. We recommend a definition such as:

Covered Data: Those newly generated or derived Scientific Data used to conduct NIH-funded or -supported research and subject to this Policy. Such data may or may not be proprietary and subject to various access controls.

3. Covered Period

AMIA Comment: We also note a need to define the expected timeframe for which grantees must steward Scientific Data. While we have numerous questions, such as, does the transfer of data to an NIH-supported or endorsed repository complete the obligation of the grantee? Will there be funding available to grantees who steward their own Scientific Data associated with tracking and satisfying data use requests? Would there be some appeal process if the volume/complexity of requests exceeds what was anticipated or funded? Or could the grantee charge some reasonable administrative fee if total costs incurred exceed some threshold?

AMIA Recommendation: We recommend the NIH address these and other questions by incorporating a concept of "Covered Period." This term would facilitate greater understanding of the obligations of grantees

Covered Period: The period of time for which the Scientific Data is expected to be maintained by the grantee and for which it is to be made available to others.

December 10, 2018

II. Purpose

AMIA Comment: This section describes what the DMSP is, but only hints at why the NIH is proposing one and how it will interact with other NIH policies. This section should describe why a DMSP is necessary and what a DMSP will achieve.

AMIA Recommendation: The draft DMSP should bolster the Purpose section by adding language similar to the introductory language, beginning, "NIH has a longstanding commitment to making the results and accomplishments of the research that it funds and conducts available to the public. Increasing access to scientific data resulting from NIH funding or support offers many benefits and reflects NIH's responsibility to maintain stewardship over taxpayer funds." AMIA recommends the draft DMSP adds to this with the following:

> *Specifically, systematic management and sharing of scientific data and results enables researchers to more vigorously test the validity of research findings, strengthen analyses by combining data sets, access hard-to-generate data, and explore new frontiers. Data management and sharing also informs future research pathways, increases the return on investment of scientific research funding, and accelerates the translation of research results into knowledge, products, and procedures to improve health and prevent disease.*
>
> *This Policy seeks to identify, adopt, and credit data management and sharing best practices, consistent with FAIR (Findable, Accessible, Interoperable, and Re-usable) data principles, so that the United States remains the leader in biomedical and life sciences research. This Policy establishes the requirements and responsibilities of researchers generating scientific data resulting from NIH-funded or -supported research and it will govern development and implementation of other NIH Policies related to the management and sharing of scientific data, such as the NIH Genomic Data Sharing Policy, the NIH Policy on the Dissemination of NIH-funded Clinical Trial Information, and the Intramural Research Program Human Data Sharing (HDS) Policy.*

III. Scope and Requirements

AMIA Comment: We applaud the NIH for considering a comprehensive DMSP that would "apply to all intramural and extramural research, funding or supported in whole or in part by NIH, that results in scientific data, regardless of NIH funding or mechanism." A pan-NIH DMSP will improve our national culture of data sharing, as well as facilitate the FAIR data principles.

We also support submission of a Data Management and Sharing Plan (Plan) as part of the funding/support application process and articulating if there are perceived barriers to sharing scientific data in this Plan. Finally, we greatly appreciate that these draft policy provisions state that "Reasonable costs associated with data management and sharing could be requested under the budget for the proposed project."

December 10, 2018

AMIA Recommendation: We urge the NIH to proceed with the proposed DMSP scope, ensuring that the policy requirements are constructed in a way that both small and large awardees can comply. While we agree that it is important for all NIH research to be subject to this policy, regardless of funding or mechanism, the policy must maintain flexibility to accommodate individual ICs and individual project characteristics.

AMIA recommends the NIH draft this section as "III. Scope" and position the aspects of the current provisions related to "requirements" in the next section, "IV Requirements for Data Management and Sharing Plans." The draft DMSP could expand on the rationale for its scope, similar to the Purpose section. We discuss issues related to IC-specific requirements and "reasonable costs," for data management and sharing below.

IV. Requirements for Data Management and Sharing Plans

a. Plan Review and Evaluation

AMIA Comment: Establishing a flexible, yet consistent, and fair review and evaluation strategy will greatly improve the likelihood that this DMSP is successful. We note that the proposed policy envisions that review and evaluation would be the primary responsibility of the funding or supporting NIH IC, "which could be implemented in a variety of ways…" and that this section delineates how various funding mechanisms might differently approach the task of review and evaluation. We are generally supportive of this strategy as long as the DMSP provides direction for ICs to rationalize and harmonize their specific requirements.

As it relates to Extramural Grants, we are concerned that scoring in a binary way has contributed to our current shortcomings in quality data management and sharing. As stated previously, we view rigorous review and evaluation of Plans as a means to improve the FAIR-ness of data and encourage the NIH to treat these Plans as scorable elements of certain grant applications.

AMIA Recommendation: We recommend the DMSP establish parity between the rigor of Plan review/evaluation and amount of NIH funding support. We strongly recommend that the draft DMSP encourage ICs to factor the quality of the Plan into the overall impact score through the peer review process for those grants that are supported at high levels or support programmatic priorities. While we support negotiation, making Plans scorable will improve the use of best practices and the general management and sharing posture of applicants far more efficiently than an "acceptable or unacceptable," evaluation schema. Rather than discouraging ICs from factoring Plan reviews/evaluations into the overall impact score, AMIA recommends ICs view quality Plans as essential to important research and design evaluation schemas to reflect this view.

Alternatively, the ICs could incentivize quality Plans by funding data management and sharing activities in an amount corresponding to the completeness of the Plan. For example, specific support of data managing and sharing activities might reflect the completeness of the plan, scored as "unsatisfactory" (0% of requested funds), "minimal" (25%), "adequate" (75%), "excellent" (100%). (Percentages for illustration only).

This recommendation notwithstanding, we do see value in considering a binary evaluation in limited circumstances, such as small grants to new investigators, or in cases where scientific data cannot be de-identified and shared.

b. Plan Elements

AMIA Comment and Recommendation: Given the extent of information expected as part of a Plan, we do not envision a 2-page limit will be sufficient in most circumstances. Rather than setting arbitrary page limits through the DMSP, we recommend the NIH leave length and depth of Plans to peer review and IC guidance.

We are generally supportive of the Plan Elements listed. However, we believe there is a need to include additional Elements so that applicants can describe their Data Management activities. We also recommend "Data Preservation and Access Timeline" be included as a sub-point of "Data Preservation and Access," rather than a standalone Element. Below we offer comment and recommendation for each of the listed Elements.

i. Data Type

AMIA Comment and Recommendation: We recommend listing the find the term "rationale" in this section confusing. Given that the DMSP clearly articulates a rationale for scientific data preservation and sharing, we recommend this section simply state:

1. Data Type: Indicate the types and estimated amount of scientific data that will result from NIH-funded or -supported research and indicate how ~~the rationale for which~~ scientific data will be preserved and shared.

1.1. Amendments: We recommend inserting "expected" following "scientific data" in 1.1 to reflect that the data actually collected may change slightly over time. The expectation should be that the Plan will be directionally correct and complete, but that it could be subject to amendment. Further, we recommend rewording the second sentence of 1.1 as follows:

Describe the data modality (e.g., imaging, genomic, mobile, patient-reported, and survey) and whether the scientific data will be individual, aggregated, or summarized, and whether the data will be ~~how~~ raw or processed ~~the data will be~~.

1.2. Amendments: We recommend adding the word "metadata" to 1.2, and we encourage the NIH to reference this defined term as appropriate throughout the document.

Describe any other information that is anticipated to be shared along with the scientific data, such as relevant associated data, and any other information necessary to interpret the data (e.g., study protocols ~~and~~ data collection instruments, and other metadata).

ii. Related Tools, Software and/or Code

December 10, 2018

AMIA Comment and Recommendation: AMIA supports efforts to make tools, software and/or code available for use, if such artifacts were developed as the result of NIH funding. However, there is a fundamental difference between sharing data and sharing code or software, particularly if the code is considered proprietary, such as a purchased stat package. The intent of this policy should be twofold: (1) To improve replicability by ensuring transparency in how data were transformed and (2) encourage the sharing of related tools, software and/or code generated through NIH funding. The intent should not be to make researchers provide an analytic environment, open source or otherwise. The use of data and workflow diagrams, which graphically depicts at a high level the data sources, operations performed on the data, and the path taken by the data through information systems and operations may be useful.

While we support the use of alternative free or open source code, we do not view the DMSP as an appropriate vehicle to encourage such solutions. The effort to identify such tools could be significant and may require skills well beyond those of the investigator and requiring assistance from staff not included in any of the grant funding. We recommend the following changes to reflect these recommendations:

ii. Related Tools, Software and/or Code: Indicate what tools, software and/or ~~computer~~ code will be used to process or analyze the scientific data ~~(the inclusion of scripts may be helpful)~~, why the software/code was chosen, and whether it is free and open source. Also indicate whether tools, software and/or code were developed to conduct NIH-supported research resulting in scientific data and if such artifacts are expected to be shared. ~~If software/code that is not free and open source is needed to access or further analyze the scientific data, briefly describe why this particular software/code is needed. Describe whether there is an alternative free and open source software/code that may be used to further analyze the scientific data.~~ The inclusion of scripts and the use of data and workflow diagrams, which graphically depicts at a high level the data sources, operations performed on the data, and the path taken by the data through information systems and operations may be useful.

iii. Standards

AMIA Comment and Recommendation: AMIA appreciates the NIH pointing towards and encouraging use of established data standards, common data elements, and other publicly funded initiatives. We support leveraging this DMSP to encourage the use of existing data standards and common data elements (CDEs) to "facilitate broader and more effective use of scientific data and to advance research across studies." We hope that, over time, researchers will coalesce around common standards when appropriate and that when common standards can be used they are used. This will only happen if Plans are critically peer reviewed by experts trained in the systematic collection, analysis, and application of data. We recommend the following changes to reflect these recommendations:

iii. Standards: Indicate what standards, if any, apply to the scientific data to be collected, including data formats, data identifiers, data models, definitions, metadata and other data documentation, including terms of use. NIH encourages the use of existing data standards, such as standards for

collecting and representing scientific data and information describing the scientific data. NIH encourages the use of common data elements (CDEs) to facilitate broader and more effective use of scientific data and to advance research across studies. For assistance in identifying NIH-supported CDEs, the NIH has established a Common Data Element Resource Portal. For a list of established clinical data standards, please see the most recent Office of the National Coordinator for Health Information Technology Standards Advisory.[3] Where commonly accepted standards don't exist, the Plan should include description of these standards in this section.

iv. Data Preservation and Access

AMIA Comment and Recommendation: AMIA encourages the NIH to be more prescriptive in its expectations that Plans leverage NIH-supported data repositories.[4] AMIA recommends the NIH incentivize the deposition of scientific data into NIH-supported data repositories by scoring or funding such Plans higher than Plans that do not use an NIH-supported or NIH-approved data repository (unless sufficient justification can be made) and by allowing Plans that leverage such repositories to forego this section of the DMSP. This may require the NIH to better understand the relative strengths and weaknesses of repositories currently/potentially supported by the NIH, but it will improve the likelihood of long-term data FAIRness. AMIA recommends the NIH develop a formal endorsement process to approve and list preferred repositories for scientific data and scientific software artifacts.

In addition, the NIH should use information gathered during the 2016 RFI on "Metrics to Assess Value of Biomedical Digital Repositories," to inform policy development in this area. While AMIA acknowledged there "will be no 'one-size fits all' scorecard" in comments to this RFI, we provided several recommendations for the NIH to develop a rating schema for deposition repositories and knowledgebases.[5]

If Plans wish to rely on data repositories other than those supported or endorsed by NIH, we recommend the following aspects be articulated (we reference existing sub-element numbers below):

4.1 Amendments: Data Deposition and Archiving: Indicate where scientific data will be archived to ensure its long-term preservation. If scientific data will be stored in an existing repository, provide the name and URL web address of the repository. If an existing repository will not be used, indicate why not and how scientific data preservation will be assured (e.g., in a newly created repository or by the investigator's organization).

4.2 Amendments Discoverability: Indicate how the scientific data will be made discoverable and whether a persistent unique identifier or other standard indexing tools will be used.

4.3 Amendments Security: Describe any provisions for maintaining the security and integrity of the scientific data (e.g., encryption and backups).

[3] https://www.healthit.gov/isa/
[4] https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
[5] AMIA Comments available at: https://www.amia.org/sites/default/files/AMIA-Response-to-NIH-RFI-on-Metrics-to-Assess-Value-of-Biomedical-Digital-Repositories.pdf

4.4 Amendments Plan Alternatives: Describe alternative plans for maintaining, preserving, and providing access to scientific data should the original Plan not be achieved.

4.5 Amendments Barriers: If perceived barriers to ~~sharing~~ preserving and making accessible scientific data exist ~~(e.g., sharing includes specific restrictions or sharing is not possible), outline how scientific data will be managed and preserved and~~ include an explanation of the perceived barriers.

4.6 Amendments Other Considerations: Indicate whether additional considerations are needed to preserve and make accessible ~~implement~~ the scientific data. ~~Plan (e.g., prior permission to use a specific repository).~~

4.7 Amendments Biospecimens: Indicate whether scientific data generated from humans or human biospecimens will be available through unrestricted (made publicly available to anyone) or restricted access (made available after the requestor has received approval to use the requested scientific data for a particular project or projects). If the scientific data will be shared through a restricted access mechanism, describe the terms of access for the data.

New 4.8 Timeline: Provide information on the anticipated timeframes for scientific data storage and accessibility, and criteria for how decisions affecting scientific data storage and accessibility will be made throughout the course of the study.

New 4.9 Amendments: Secondary Use Timeline: Describe when the scientific data will be made available to secondary data users. This should be expressed in relation to some critical event, such as the publication of the major study findings, the end of data collection, or other similar activity.

v. Data Preservation and Access Timeline

AMIA Comment and Recommendation: AMIA recommends the DMSP merge Element 5 as subordinate points of Element 4 (see above Elements 4.8 and 4.9). We recommend that Element 5.2 be removed from the DMSP.

vi. Data Sharing Agreements, Licensing, and Intellectual Property

AMIA Comment and Recommendation: AMIA supports the expectation that scientific data will be broadly available, consistent with privacy, security, informed consent, and proprietary issues. We note that this information may be duplicative with information provided in prior Elements, such as barriers to preservation / access, and we encourage NIH to reduce sections that overlap in intent or required content.

6.1 Amendments Data Sharing Agreements: Describe any existing data sharing agreement(s), outlining the responsibilities of each party, as well as how scientific data can and cannot be used.

6.2 Amendments Licensing: Describe any existing ~~general~~ licensing terms, and any limitations on the scientific data use and reuse based on these terms. Describe whether the licensing is imposed by the applicant institution or whether it comes from any existing agreement(s).

6.3 Amendments Intellectual Property: If applicable, indicate how intellectual property, including invention or other proprietary rights, will be managed in a way to maximize sharing of scientific data. Include any information relevant to the intellectual property rights associated with the scientific data, such as whether the intellectual property stems from an existing agreement or is anticipated to arise from the proposed research project itself.

vii. Oversight of Data Management

AMIA Comment and Recommendation: AMIA recommends removal of this section, given that grantees already provide personnel information in other parts of the grant. If it remains in the draft DMSP, we recommend a focus on the role rather than the individual to describe data management oversight and execution of the Plan.

viii. Other Considerations

AMIA Comment and Recommendation: AMIA views the additional considerations as important context that could be used to

V. Compliance and Enforcement

AMIA Comment and Recommendation: AMIA generally supports the compliance section "During the Funding or Support Period," and "Post-Funding or Support Period." However, we note that data management is an ongoing process and that a management plan is updated, modified, and versioned. We anticipate that this part of the Plan could be part of the progress report statement. As for data sharing, we reiterate our recommendation that NIH develop a formal endorsement process of preferred databases and knowledgebases. These endorsed repositories would facilitate DMSP compliance and enforcement by having transparent terms and conditions and abide community consensus best practices. Researchers who use these NIH endorsed repositories would have a streamlined compliance process.