January 16, 2017

Dr. Joe V. Selby, MD, MPH
Executive Director
Patient-Centered Outcomes Research Institute
1828 L Street, NW, Suite 900
Washington, DC 20036
Submitted electronically at: http://www.pcori.org/webform/data-access-and-data-sharing-policy-public-comment

Re: Request for Comment: Data Access and Data Sharing Policy

Dear Dr. Selby:

The American Medical Informatics Association (AMIA) appreciates the opportunity to submit comments regarding Patient-Centered Outcomes Research Institute's (PCORI) request for comment on its Data Access and Data Sharing Policy.

AMIA is the professional home for more than 5,000 informatics professionals, representing researchers, front-line clinicians, public health experts, and educators who bring meaning to data, manage information and generate new knowledge across the health and research enterprise. Numerous AMIA members are PCORI-funded investigators, and as such, we have keen interest in this policy.

AMIA enthusiastically supports development of policies for data access and sharing. We share in PCORI's principles of open science by maximizing the utility and usability of data collected in research projects that it funds. Recently, AMIA published the first in a series of Policy Principles & Positions.[1] Among them was an articulation of our belief that data sharing among researchers is "foundational to advance scientific discovery; improve benefit / risk assessments; conduct comparative effectiveness research; prevent medical errors; and promote biomedical research rigor, transparency and reproducibility."[2] Further, we have developed an initial set of Policy Positions for Data Sharing in Research, which enumerate several activities AMIA supports based on our Policy Principles. Below, we highlight key recommendations for PCORI to consider as they finalize this data sharing and access policy.

[1] "AMIA Public Policy Principles and Policy Positions, 2016 – 2017," available at http://bit.ly/2gPB52N
[2] Ibid., Data Sharing in Research Policy Principle (pg. 10)

January 16, 2017

Key AMIA Recommendations

In order to improve its data sharing policy, PCORI should consider:

- Requiring preliminary data sharing plans as part of award applications, not just as a post-reward requirement;
- Earmarking specified amounts of grant funding for data preparation and curation, in addition to covering reasonable costs associated with maintaining and depositing the Full Data Package; (including de-identification);
- Defining metadata, including requisite attributes and vocabularies used to provide attributes, as an explicit component of the *Analyzable Data Set*, and encourage the use of emerging community metadata models; and
- Establish ways to ensure data originators receive credit for their work

Below we outline our recommendations in more detail, and we address PCORI's specific questions related to this draft policy. We hope our comments are helpful as you undertake this important work. AMIA would gladly welcome representation from PCORI at the 2017 iHealth 2017 Clinical Informatics Conference, May 2 - 4, 2017, as part of the ongoing dialogue we hope to foster by responding to this RFI. 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.

Sincerely,

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

Thomas H. Payne, MD, FACP, FACMI
AMIA Board Chair
Medical Director, IT Services, UW Medicine
University of Washington

Enclosed - Detailed AMIA Recommendations and Comments to PCORI Questions

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

January 16, 2017

Detailed AMIA Recommendations and Comments to PCORI Questions

Data sharing plans as part of award applications
AMIA recommends that a preliminary data sharing plan be a required component of all PCORI funding applications. Data sharing has become such an important proximal output of research that we believe the relative value of a proposed project should include consideration of how its data will be shared. Requiring a pre-award description of intended data sharing of federally-funded research could address fundamental deficiencies in biomedical and clinical research. We commend PCORI for requiring a data sharing plan post-award, but we believe PCORI could play a leading role in helping improve the data sharing stance of clinical research by asking investigators to think about data sharing in the earliest stages of design.

As part of Section B "Awardee Requirements" AMIA recommends PCORI adds to 1.a. more detail to the required data sharing plans as follows:

a. Preparation of the Full Data Package in accordance with a documented data management plan that describes the data one expects to acquire or generate during the course of the research project, as well as how the data will be organized, managed, stored, and preserved. This should include:
    1. The proposed repository for the data and its accompanying data dictionary;
    2. The metadata standards used to create the data dictionary; and
    3. The plan, if any, to update datasets after they have been deposited in a repository.

Funding for data preparation and curation
AMIA fully supports and appreciates PCORI's intention to provide funding for costs associated with maintaining and depositing the Full Data Package in a suggested repository, as outlined in Section D of the data sharing plan. However, evidence suggests that dedicated funding for data curation and preparation for sharing is an important contributor to the sustainability of research ecosystems.[3] [4] By including data sharing costs in proposed budgets, and by providing guidance regarding best practices, PCORI creates incentives to share, collaborate and advance data sharing capabilities. This would further underscore the need for quality data sharing plans to be considered as part of the application process.

Metadata
Metadata are essential to understanding all datasets, especially in the context of data sharing in biomedical research. We appreciate PCORI defining the Full Data Package by including language

[3] National Academy of Medicine (formerly Institute of Medicine) "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk," Jan. 2015 http://bit.ly/1Vwtnbi
[4] Borne, P., Lorsch, J., Green, E., "Perspective: Sustaining the big-data ecosystem," *Nature*. November 2015. 527, S16–S17

American Medical Informatics Association | AMIA
4720 Montgomery Lane, Suite 500 | Bethesda, Maryland 20814

borrowed from the National Academy of Medicine to describe metadata.[5] However, we recommend PCORI explicitly define and/or include mention of metadata in the Full Data Package definition. Should PCORI seek a specific definition for metadata, we recommend defining metadata as "data about data and all the processes that produce, streamline and output data."[6] Specifically, this would include attributes describing the data and specification of allowable values of those attributes, including accepted vocabularies and terminologies describing key biomedical concepts. Emerging community metadata models could be useful towards improved, more consistent use of metadata[7]and should be promoted as standards for data sharing plans.

Support for data originators

Another important contributor to the sustainability of research ecosystems is ensuring that those who create and/or contribute to public datasets and software that others find useful are credited.[8] AMIA recommends PCORI establish ways to ensure data originators receive credit for their work as part of the process for Third-party Requests.

	PCORI Question	AMIA Comment
Q1.	What should be the retention period for how long PCORI research awardees must retain the full data package for sharing?	AMIA believes seven years is appropriate. We support the policy as written.
Q2.	What restrictions on use of PCORI research awardee data are important to include in the data use agreement executed by third parties requesting access to the data (e.g., data will not be re-identified, data will be used only for research and not commercial purposes, data will be used only for research in the same therapeutic area as the original research project through which the data were collected)?	AMIA recommends PCORI require attribution for use of awardee data, in addition to the prohibition on re-identification of PCORI awardee data. Beyond those basic requirements we do not believe it would be feasible, practical, nor would it be in keeping with the spirit of open data, to restrict the use of awardee data by the public. A more restrictive access policy would inhibit reuse along the distribution chain, including for manufacturers, PBMs, health insurance companies, from accessing PCORI data even for drugs or devices they manufacture, distribute or insure.

[5] "Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk" defines metadata as "data about the data metadata (e.g., protocol, statistical analysis plan, and analytic code)"

[6] M. lzzo, Biomedical Research and Integrated Biobanking: An Innovative Paradigm for Heterogeneous Data Management, Springer Theses, DOI 10.1007/978-3-319-31241-5_2

[7] Examples includ the MiXXX models or the W3C HCLS dataset model

[8] Piwowar, H., Vision, T., "Data reuse and the open data citation advantage," *Peer J.* 2013. 1:e175

Q3.	What qualifications/credentials should be required for a third party requesting access to PCORI research awardee data (e.g., education level, specific scientific expertise)?	AMIA does not support the specification of qualifications / credentials for third party requesters. We believe anyone should be able to use open data.
Q4.	What documentation should a third party requester be required to provide when applying to access PCORI research awardee data (e.g., research question and lay protocol summary, full protocol and statistical analysis plan, institutional review board approval)?	In a scenario of strictly de-identified data, third parties should provide a research question, at a minimum, and proposed analysis plan to access PCORI research awardee data. If the data is used for exploratory purposes, a research question should suffice. Should the data be identifiable, third parties should agree to standards / protocols commensurate with IRB approval.
Q5.	Should PCORI establish data repository standards (e.g. security, curation standards)? If so, what should the minimum standards be for a repository to qualify as a PCORI suggested repository for PCORI research awardee data?	AMIA strongly supports the use of data repository standards and we recommend PCORI seek to develop standards that are consistent with other federal funders, including any models that might be supported or approved by the NIH, and with practices of existing data repositories and indexing effort.[9] We strongly caution PCORI against reinventing wheels as a function of its pilot project.
Q6.	What is the appropriate model for informed consent that should be included in the policy?	Applicants for PCORI funding should be required to include detailed plans for obtaining consent for data reuse as part of the data sharing plans submitted with each application. PCORI could make a significant contribution to the research community by preparing standard text for this aspect of informed consent.
Q7.	Do you have any other comments? Specifically, we seek comment about any of the following sections: (I) Purpose, (II) Applicability, (III) Definitions, and (IV) Policy.	See Key and Detailed Recommendations.

[9] AHA Approved Data Repositories: https://professional.heart.org/professional/ResearchPrograms/UCM_461443_AHA-Approved-Data-Repositories.jsp