Compiled Public Comments on Request for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data

Guide Notice Number: NOT-OD-22-131

May 12, 2022 – June 27, 2022

Table of Contents

ID: 2008

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Type of Organization: Biotech/Pharmaceutical Company

Role: Scientific researcher

Comments:

Data sharing agreements between institutes can result in lengthy negotiations, become a big burden on PIs, and cause major delays in projects (sometimes years). It would be most effective if NIH could share a sample/template/best-practice data sharing agreement that can be adjusted by collaborating institutes for use in the context of NIH programs. Starting from such a neutral, NIH agreement template would greatly facilitate the negotiations.

Email: brudeman@gmail.com

ID: 2009

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Type of Organization: Nonprofit Research Organization

Role: Scientific researcher

Comments:

I am worried that having the public with no research background will misinterpret and misuse NIH funded studies. In order for the public to best make use of our research data, they need to be more literate in the process of science and analyses of data. Therefore, I suggest a system be created to enhance the education of the public on data use, interpretation, and methods used to obtain the data before data are allowed to be shared. I know this is asking a lot. However, I worry that with so many possible interpretations of the meaning of the results obtained from data, it will become even harder to reach consensus on best practices.

ID: 2010

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Name: R. Katherine

Name of Organization: Fox Chase Cancer Center

Type of Organization: Health Care Delivery Organization

Role: Scientific researcher

Comments:

Protecting a patient's identity is important. All specimens, reports, pathology etc. should only contain a patient's initials, study number assigned to then at study registration and a date of birth (DOB is not PHI)

Description: feedback survey on protection of patient identity

ID: 2011

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Type of Organization: University

Role: Scientific researcher

Comments:

If implemented these new regulations will destroy research which might ultimately benefit the patient. The added bureaucracy will stop research happening. Please do not go through with them.

ID: 2012

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Name: Carl

Name of Organization: Johns Hopkins Bloomberg School of Public Health

Type of Organization: University

Role: Scientific researcher

Comments:

I have found that institutional agreements to share data between institutions can be very cumbersome and time consuming with lawyers from both institution becoming involved in the developing and executing of users agreements. It would be much easier to have NIH standardized agreements approved a priori. Moreover, it is cumbersome and time consuming making data sets available. NIH should develop a set of macros and standardized labeling that can quickly be applied to data sets as well as provide recourses and guidelines for organizations to have cloud storage for data sets.

ID: 2013

Submit date: 5/17/22

I am responding to this RFI: On behalf of myself

Type of Organization: University

Role: Scientific researcher

Comments:

While it is laudable for the NIH to take personal privacy very seriously, the downstream consequences need to be considered carefully. Protection of personal health information under current standards is very high. In fact, before more restrictive rule changes be initiated, the NIH and other agencies should document clearly where the current standards are falling short. Can they document harms under the current system? If so, what are the harms and will changing rules shore up weak points in the current guidelines to make things safe? There are several facts to consider in protecting individuals: 1) New rules almost invariably lead to greater administrative burden on researchers and their institutions. This leads to increased time away from science, and higher rates of burnout among scientists. This is likely to threaten the US scientific pre-eminence in STEM since fewer talented individuals will want to go into it. 2) New rules or a change in rules governing consenting is expensive. If the rules are more restrictive, many patients already participating in studies will need to be re-consented. Going forward, more restrictive consenting procedures, need for special encryption of PHI etc. will all require significant capital to build these new systems. Who will pay for this? Consenting human subjects for research is already woefully underfunded and these rules will dampen human subjects research. 3) A large segment of the populous gives away personal information with every internet search and every online agreement they sign with tech companies. In many case, the information that is harvested is higher risk and more damaging than PHI and yet is given away freely. Is there a common model that could be pursued? Can a system be designed that allows use of human samples (for example) with limited consent and liability for researchers since the samples are being used for public benefit is the vast majority of cases, while PHI harvested by big tech is invariably used for profit? There seems to be a moral difference between the goals of science and the goals of industry, and yet it is the scientists who are shouldered with much greater administrative burden. 4) Given the increased expense of changes to protection of PHI, greater protections and safeguards will be unduly burdensome to institutions of higher learning with few rea=sources. Rich institutions will make it work, and this will put them in better position for competitive funding. Poorer institutions will be handicapped. This, unfortunately will particularly affect institutions attended by underrepresented minorities and individuals with lower incomes. The rich will get richer, while the less resourced institutions will suffer.

ID: 2014

Submit date: 5/18/22

I am responding to this RFI: On behalf of myself

Name: Gregory Simon

Name of Organization: Kaiser Permanente Washington Health Research Institute

Type of Organization: Nonprofit Research Organization

Role: Scientific researcher

Comments:

It would be helpful for recommendations to distinguish between data collected or created with explicit informed consent and data used via waiver of consent (such as data extracted from health system records or insurance claims). When considering risk of re-identification for data collected with explicit consent, investigators and IRBs may fall back on what was explained and agreed to in that consent process. When considering risk of re-identification for data used under waiver of consent, investigators and IRBs must often apply a stricter standard - since no consent was given for original use of data, much less for sharing.

ID: 2015

Submit date: 5/18/22

I am responding to this RFI: On behalf of myself

Name: Aimee Payne

Name of Organization: University of Pennsylvania

Type of Organization: University

Role: Scientific researcher

Comments:

Single cell RNAseq and other single cell analyses currently fall under the NIH GDS, and hence requires prospective informed consent. I would recommend that single cell analyses be removed from the genomic data sharing requirements if performed on fewer than 100 individuals. The rationale is as follows: 1) Valuable human tissues that are obtained for clinical purposes, otherwise discarded, and de-identified (ie, otherwise human subjects research exempt) could not be used for pilot single cell studies, because by sweeping these types of studies under the GDS, prospective informed consent would be required. 2) The NIH Genomic Data Sharing policy has historically excluded studies of bulk RNAseq unless if performed on greater than 100 individuals (plus other scenarios as outlined: https://www.niaid.nih.gov/research/when-gds-policy-applies). Giving a concrete example, if an otherwise discarded blood or tissue sample were used for bulk RNAseq, it would not fall under the GDS and would not require prospetive informed consent. But if you took the same sample, discarded the vast majority of the sample and analyzed 10,000 cells by scRNAseq, it would now fall under the GDS even though there is no increased risk of patient identification compared to other approaches that do not fall under the GDS. 3) An alternative solution is to require data sharing but to not make that intrinsically linked to the requirement for informed consent so that tissues falling under the common rule could continue to be used for research purposes.

ID: 2016

Submit date: 5/19/22

I am responding to this RFI: On behalf of myself

Name: Sheryl L. Chatfield

Name of Organization: Kent State University

Type of Organization: University

Role: Scientific researcher

Comments:

As an individual who actively generates (and reuses) qualitative and mixed methods data, I recommend some enhancements be made to this information to more particularly address the unique issues of, for example, interview data. I should say I am highly in favor of archiving and re-use of data from individual or group interviews, observation field notes and other data types traditionally associated with qualitative and mixed methods research endeavors. For example, the phrase "data and biospecimens" is general and accurate but may not be readily recognized by people who are about to participate in an individual interview. Maybe offering an alternative phrase "responses you provide during an individual or group interview, or notes researchers make while interacting with or observing you," or a footnote to show that this may be what "data" means in a given context, would be more informative (and less alarming) for some participants, especially those who have historically been underrepresented in research and may be suspicious when things like "biospecimens" are emphasized. The other issue where qualitative researchers, and their participants would benefit from more guidance includes de-identification as a process. Sometimes it is not necessary to have a name or demographic-type information to know who someone is - they may be recognizable from an experience, an organization, contact with another individual (identified by role if not name), or a location, alone or in combination. Since qualitative researchers often focus on the unique and the particular - those who might be consider "outliers" in frequentist quantitative methods - and location of participants regularly has some association with location of researchers, reasoning out who a participant is may be very possible from contextual information alone. When I was a graduate student, I was trained to remove location identifiers (e.g., substitute CITY for name of a place, PERSON for other individuals, etc.) but this is not a regular practice I have seen emphasized by multiple instructors and textbooks. And sometimes it doesn't matter, if interviews relate to a key event, or it may dilute usefulness of interview data to remove too much of the context. So I recommend some thought be given to enhancing these guidelines to at least consider or refer to some examples. Although there is an assumption that most of the data of concern addressed by these processes quantitative (including categorical), I believe an increasing proportion of scientific researchers also value unstructured data, which may include traditional types of qualitative data, to supplement their data gathered in large, lengthy and costly projects. So I think that enhancing this guidance to take multiple types of data into consideration, during the draft phase, is warranted.

ID: 2023

Submit date: 5/27/22

I am responding to this RFI: On behalf of myself

Name: Ted Henderson Senior

Name of Organization: HCW, C.A., LLC, & CWH DUBAI TECHNOLOGIES INTERNATIONAL

Type of Organization: Professional Org/Association

Role: Patient advocate

Comments:

Description: E-Acute Emergency Health-Care

Email: tjhsr6030@gmail.com

ID: 2025

Submit date: 5/31/22

I am responding to this RFI: On behalf of myself

Type of Organization: University

Role: Scientific researcher

Comments:

Proteomics appears noticeably absent as a requirement for publication, in contrast to DNA-seq and RNA-seq approaches. While databases of peptide hits etc exist, alternative proteomics approaches are gaining use in the literature (eg, aptamer-based proteomics), necessitating wording of the expectation that proteomics studies should be published in parallel with a repository deposition. The most important aspect of this for accessibility to the general scientific public is a requirement for gene/protein-product level processed data, not just the raw data. Thank you for your efforts to encourage the sharing of Big and little Data.

ID: 2026

Submit date: 6/1/2022

I am responding to this RFI: On behalf of an organization

Name: Julia Schaletzky

Name of Organization: Center of Emerging and Neglected Diseases, UC Berkeley

Type of Organization: Nonprofit Research Organization

Role: Bioethicist

Comments:

Please ensure an exception to privacy is clearly in place for sharing data (i.e. diagnostic data from a research study) with the actual participants in the study. This is currently almost never possible. For example, at our institution Covid surveillance testing was conducted and positives were identified (early in the pandemic when testing was not available to the public) but it was not possible to tell participants that their sample was positive. This can happen if samples are deidentified right away (and one cannot tell which patient it came from) but also because there is significant uncertainty around what is allowable under privacy rules among researchers who are not legal experts. It is deeply unethical to use patient protection rules as a weapon against patients, depriving them of their right to learn about their own health and take measures to prevent harm or infect others. When establishing guidelines, please clearly distinguish between rights of patients to their own data which should be always guaranteed, and rights of other parties to data. There should be no question that an investigator can share data with the patient that helped generate it. This should not require IRB approval or any other administrative steps that could be barriers to timely sharing of health information. This would also help encourage participation in research studies.

ID: 2027

Submit date: 6/6/2022

I am responding to this RFI: On behalf of an organization

Name: Shanda Hunt

Name of Organization: Research Data Services Team at University of Minnesota Libraries

Type of Organization: University

Role: Scientific researcher

Comments:

DRAFT Operational Principles for Protecting Participant Privacy When Sharing Scientific Data --We appreciate that the first two points made in this draft proposal emphasize informed consent and ensuring prioritization of the participant agreement. We would like to see this ethical consideration woven throughout the proposal. Also, perhaps expand informed consent to “any type of participant agreement” given that not all data being shared is reviewed by an Institutional Review Board (IRB). --This statement is critical: “...prioritize clarity regarding future sharing and use of scientific data.” As data repository curators, we prefer to see “data” explicitly addressed in participant agreements rather than the commonly used term “records.” --Because data repositories differ in the protections and restrictions offered, we would like to see researchers encouraged to base their descriptions of data restrictions on the repository in which they will share the data. It can be expensive and time consuming to find a repository (or re-consent participants) to fit custom restrictions. --We encourage the NIH to emphasize the properties of the data/population that make sharing the information risky, rather than regulatory determinations. Even data that do not meet "human participant research" requirements of the Common Rule may be inappropriate to share without some effort to protect it. For example, administrative data, such as campus wifi activity, can be highly identifiable and potentially sensitive, but would be described as "not human research data" according to the Common Rule. --Restricted data sharing is expensive and effortful. This policy seems to be encouraging this route rather than publicly sharing. In order to make sure research data is accessible over time, there needs to be funding for repositories to continue to mediate this access. --We would like to see additional guidance about what information and data researchers should retain when submitting to a repository. For example, if data are de-identified, how long should the original, identifying data be retained? Should researchers include or share documentation around the de-identification techniques used (e.g., which variables were re-coded or removed, local suppression reports, perturbations used if at all) with their data submission? --The automatic issuing of Certificates of Confidentiality (CoC) for all studies regardless of sensitivity puts a barrier on - and is potentially at odds with - data sharing. Our suggestion is to allow principal investigators to opt out of this, or add more granularity on the content of the data when NIH issues a CoC. --It is unclear who is responsible for the institutional review of data sharing. While the IRB may be an obvious choice, these issues require expertise outside of regulatory guidance. We recommend including or guiding researchers to institutional expertise from the libraries, data repositories, data curators, and data governance to ensure institutional and repository policies are not being violated. --We would like to see an example of when NIH would deem it acceptable to not share the research data. DRAFT Best Practices for Protecting Participant Privacy When Sharing Scientific Data --We appreciate the

coverage of both de-identification and data use agreements (DUAs) in protecting participant data, as they work together to protect privacy. We also appreciate that the guidance for the data use agreements is largely left up to the repository, with broad points to include. --Consider referencing other de-identification guidance, such as those provided by ICPSR https://www.icpsr.umich.edu/web/pages/datamanagement/confidentiality/index.html, that provide more nuanced descriptions of dealing with indirect identifiers (e.g., everything else that would fall into the HIPAA Safe Harbor #19 Any other unique identifying number, characteristic, or code). --Encourage documentation of the de-identification techniques used on shared data. This allows future users to understand the full original data, any derivatives or eliminated variables, and if it is possible to request the original data as needed. --We recommend including guidance on how long original data should be retained if the shared and preserved data were de-identified. --We recommend including or guiding researchers to institutional expertise from the libraries, data repositories, data curators, and data governance in the review of data sharing (in addition to the IRB). We recommend that if IRBs are to be the consultative bodies for data sharing issues, they receive additional training on appropriate and acceptable data sharing practices. --We would encourage NIH to broaden the description of Data Use Agreements to include click through or terms of use agreements for more open access repositories. The guidance given is applicable to these situations but is currently written to suggest only mediated DUAs are advised (which require a high administrative burden). --Again, we want to emphasize that a CoC for all human participant research seems overly restrictive and may be at odds with data sharing best practices. DRAFT Points to Consider for Designating Scientific Data for Controlled Access --We are concerned this guidance is written in a way that assumes researchers will be starting from the point of restricted sharing. We would like to see a statement encouraging researchers to share data as openly as possible, while protecting the privacy of participants. Should the default be without access controls, unless the list elements apply? If not, then perhaps this should be reframed to be when NOT to consider access controls. We would like to reiterate that restricted sharing is more expensive to maintain. --In order to have participants consent to sharing without restriction, researchers would have already needed to plan for this to write the consent forms in this way. The default for many IRB consent templates is more restricted. --To reason #3, add in that de-identification reduces the analytical or reproducible utility of the data.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/kdxNUWFzxT.pdf

Description: Response to NIH RFI on Protecting Privacy - UMN UL RDS

Email: hunt0081@umn.edu

ID: 2033

Submit date: 06/16/22

I am responding to this RFI: On behalf of myself

Name: Tim Kinkead

Name of Organization: Preva Group, LLC

Type of Organization: Other

Role: Member of the Public

Comments:

The following feedback is provided in response to notice number NOT-OD-22-131. The operating principles adequately address respect for and protection of patient privacy. The following suggestions may strengthen these principles: 1) The fourth principle regarding institution review is unclear. The concept of an accountable institution that must respect and protect privacy would strengthen privacy. Citizen research could continue within a digital research environment managed by trusted institutions. 2) Participant privacy should be respected and protected regardless of how the data is collected. Collection of data from non-traditional research settings should have the same protections and consent requirements as any other collection method. The current wording of "warrant strict privacy considerations" may indicate the other principles do not apply when they should. 3) The purpose of sharing data for subsequent research is to benefit the participant's community. The words "There may be justifiable exceptions to sharing scientific data" are too vague. Researchers seeking to horde participant data for personal gain likely view their rationale to not share as justifiable. The exceptions should be rare and only allowed when the risks to the participants outweigh the benefits to their community.

ID: 2034

Submit date: 06/17/22

I am responding to this RFI: On behalf of myself

Name: Melissa Haendel

Name of Organization: University of Colorado

Type of Organization: University

Role: Scientific researcher

Comments:

These are very important considerations and the new supplemental information has not yet been widely publicized. The request for additional comments is coming at a time when many folks are attending graduations or are on vacation. I would highly recommend you keep the comment period open longer :-). Will write a real response later.

ID: 2035

Submit date: 06/17/22

I am responding to this RFI: On behalf of myself

Name: Jen Collinger

Type of Organization: University

Role: Scientific researcher

Comments:

For research projects with small numbers of subjects, the possibility of identification remains high. Controlled-access repositories are a great approach to minimize the risk of breach of confidentiality, but I am interested to hear if there are additional recommendations. How can this risk be best described in consent forms for small-n studies?

Email: collinger@pitt.edu

ID: 2038

Submit date: 06/20/22

I am responding to this RFI: On behalf of an organization

Comments:

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/efXQAJNkhu.pdf

Description: Comments on Draft Supplemental Information on NIH DMS Policy-Privacy Protections

Email: spanicker@primr.org

ID: 2041

Submit date: 6/23/22

I am responding to this RFI: On behalf of an organization

Name: Dr. Emma Meagher and Lauren Steinfeld

Name of Organization: Penn Medicine

Type of Organization: University

Role: Institutional official

Comments:

Attached please find response from Penn Medicine.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/MdodQXaulT.pdf

Description: Penn Medicine Response

Email: kristen.molloy@pennmedicine.upenn.edu

ID: 2044

Submit date: 6/24/22

I am responding to this RFI: On behalf of myself

Type of Organization: Other

Role: Member of the Public

Comments:

Comments on the DRAFT Best Practices for Protecting Participant Privacy When Sharing Scientific Data:
1. Ensure Appropriate De-identification. The DRAFT does not provide clear principles and best practices for de-identification. It does not assist covered entities to understand what is de-identification, and the options available for performing de-identification. The de-identification approach used by the National Institute of Mental Health Data Archive (NDA) is to create an NDA Global Unique Identifier (GUID) based on a participant's PII (first name, middle name, last name, sex, date of birth, and city/municipality of birth). While PII will not be submitted to NDA, the GUID hash codes allow NDA users to link participant data records across different studies and across time and locations. In other words, the GUID becomes an identifier for linking individual-level data. Is this the NIH's recommended best practice for de-identification?

ID: 2045

Submit date: 6/24/22

I am responding to this RFI: On behalf of an organization

Name: Stacey Berg

Name of Organization: Baylor College of Medicine

Type of Organization: University

Role: Institutional official

Comments:

Thank you for the opportunity to comment on this guidance. Our comments are attached.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/ffLXHlAGPi.pdf

Description: Baylor College of Medicine comments

ID: 2046

Submit date: 6/24/22

I am responding to this RFI: On behalf of an organization

Name: Kristin West

Name of Organization: COGR (Council on Governmental Relations)

Type of Organization: Other

Type of Organization-Other: Nonprofit association of research universities

Role: Institutional official

Comments:

Please see the attached letter with our comments.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/TbDOokaCgs.pdf

Description: Response to NIH NOT OD 22 131

Email: kwest@cogr.edu

ID: 2055

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Kimberly Morales

Name of Organization: International Society for Biological and Environmental Repositories

Type of Organization: Nonprofit Research Organization

Role: Member of the Public

Comments:

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/ZRMTUOsPYh.pdf

Email: info@isber.org

ID: 2056

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: David Gordon

Name of Organization: Palantir Technologies Inc.

Type of Organization: Other

Type of Organization-Other: Software Provider

Role: Institutional official

Comments:

Introduction We are pleased to have the opportunity to provide our comments on the Draft Supplemental Information to the National Institutes of Health (NIH) Policy for Data Management and Sharing (DMS Policy) to address privacy considerations. We have deep and extensive experience supporting the government in leveraging data across organizational boundaries, while simultaneously enabling high quality data governance and adhering to data protection laws, and wider privacy and civil liberties norms. In this submission, we seek to offer our high-level perspectives on recommended additions to the supplemental information developed by NIH for the DMS Policy. Our recommendations focus on the “Draft Best Practices for Protecting Participant Privacy When Sharing Scientific Data”. Specifically, we recommend adding two additional strategies under the first best practice, “Ensure Appropriate De-identification” in order to improve the robustness of the subsequent privacy framework. Recommended Additions - "Because combining new data with existing datasets will increase the risk of data re-identifiability, there must be a process for evaluating risk and approving the use of each additional dataset in tandem with the original dataset, even if the new data is itself unrestricted or publicly available data. Data owners should implement data access restrictions, to limit how data is used." An anonymized dataset that might seem safe for wider sharing can still be susceptible to re-identification—even if it’s been verified to not reveal individual information. Partial information from the anonymized data, linked with insights from other data sources (e.g., datasets already in the public domain, additional uploaded datasets) can be used to reverse engineer aspects of the anonymization process. To further protect data from re-identification, we recommend that NIH implements data governance policies to track and restrict data access, and ensure data users only have access to PHI/PII data when strictly necessary. Data asset access should be tied to a specific purpose. Data owners should govern how distinct groups of users interact with the same data based on classification, role, attribute, or purpose, configuring multi-tier access controls at the row, dataset, and project level. After data is uploaded and scanned for protection validation, all subsequent processing should rely on purpose-based access controls, which requires users to first state their purpose before being granted access. This means that instead of getting access to individual datasets, users get access to purposes. Purposes are set by data governance teams or data owners to contain data specifically scoped to help users meet their goals—no more, no less. Purposes make it clear why a user might have access to some sensitive data, and when this access is no longer necessary. With purpose-based access controls, sensitive data will have an additional layer of security that helps researchers use data judiciously. - "Because data de-

identification does not always succeed, it is necessary to establish a standard operating procedure (SOP) for the removal of all PHI/PII (or other data upon request) from the dataset, as well as from all derived data and findings. Researchers and data owners must be able to track the data lineage of all computational outputs to successfully correct all datasets and records with PHI/PII data." Sometimes, de-identification does not succeed, and PHI/PII is unintentionally retained when uploaded and used by researchers to derive further data and findings. When the PHI/PII is discovered, if there is no contingency mechanism in place for transparency, the only way to ensure it is removed is to completely lock or delete all related research projects, potentially losing months of research and findings. Establishing an SOP that details a removal process of PHI/PII from not only the original dataset but also all derived data is crucial to ensuring research projects can remain uninterrupted and minimally affected, even through a PHI/PII removal process. To successfully amend and correct datasets and records, it is necessary to automatically maintain a complete, comprehensive history of all data and its schema, including an archive or all raw information from legacy systems. By maintaining a record of the information/logic used to produce outputs, data lineage can always be tracked and data quality issues can be proactively fixed. This full visibility into data lineage history ensures that data providers, data owners, and researchers can trust the data being used, understand any given changes, and make informed decisions with confidence.

Email: lchu@palantir.com

ID: 2057

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Joanna Groden

Name of Organization: University of Illinois Chicago

Type of Organization: University

Role: Institutional official

Comments:

We strongly support research data sharing as a mechanism to advance knowledge and are pleased to learn that NIH is also focusing on the challenges of data reidentification following implementation of its new Data Management and Sharing Policy. We would like to suggest some potential considerations for finalizing the draft documents to increase protections for human participants who generously participate in research. Current De-Identification Challenges De-identification law and policy do not currently address our technological ability to re-identify datasets as they are compiled or overlaid. With continued advances in machine learning, natural language processing, data science and artificial intelligence, we must acknowledge that researchers and other users of data will re-identify datasets, deliberately or inadvertently. This creates a tension between sharing and trust for the curation of public data and has significant potential for downstream harm of minoritized individuals and groups. From a computer science perspective, the question of differential privacy and data fidelity further arise as complications for the described intent of maximal data sharing under the new Data Management and Sharing Policy. While some options are available to obfuscate data fully to enable human subject protection, anything that should not be re-identified or which will be shared broadly will have to be of incredibly low quality. This additionally then greatly limits the utility and any benefits possible for either researchers or members of the public. We recommend ongoing policy review and updating, as well as funding for education and tools to allow enhanced tracking and notification of dataset reuse, improved understanding of re-identification harm, increased development of techniques for reuse which respect privacy and security, and greater obligations for data protection by those seeking to reuse data. Tracking and notification would allow both alerts where datasets are being used, providing credit and impact to the original dataset generator, and to identify potential participant harm. Certificates of Confidentiality The draft guidance includes the line “Protections afforded by Certificates apply to all copies of a dataset in perpetuity.” We recommend that this statement be reviewed and clarified as current Certificates of Confidentiality only deal with directly identifiable data. Current policy allows for more expansive sharing of de-identified data without further review from IRBs. It is unclear if this sentence, provided here only in guidance, would change current policy and practice. Further, a definition is needed of what comprises “copies of a dataset”. Copies could include fully identifiable data, de-identified data or other derivative copies to significantly change approaches to longitudinal or meta-analysis research. If all potential identifiable derivative copies must be reviewed by an institutional review board, academic institutions will need additional guidance on what to consider related to these derivative datasets. Reuse by Private Partners and Secondary Users Other considerations for data reuse and sharing are the increased collaborations between academic institutions and private partners, whether for drug development and

clinical trials, or other research. Further, as researchers seek broad sharing of their data through approved repositories, clarity is needed as to obligations and responsibilities for the primary data generator and depositor, and for any secondary users, should a Certificate of Confidentiality (if those are expanded) or other protection requirements are violated. We recommend that additional guidance and directions be created which outlines how future use limitation may be monitored, particularly with private partners and secondary users. Mediated Reuse We appreciate the importance of prioritizing access to human participants for trials where there may be direct benefit, but are concerned about the limitations of reproducibility should significant numbers of participants opt out of broader data sharing. Draft Principle 3 may override the ability of researchers to meet needs for reproducibility in de-identified data sharing. We recommend identification of techniques that allow for an interim level of sharing with greater investigator control. Rarity of Non-Traditional Research Draft Principles 5 and 6 include specific language used regarding non-traditional research and surveillance. While we recognize and advocate for the need to protect individuals in these instances, we do not see a need to require stricter privacy controls or documentation of consent for subsequent use when such a consent process would be the only mechanism that creates an identifiable record. For public health and health services research, this could have significant negative impact, risking additional disclosure, and potentially harming or reducing the ability to appropriately develop effective methods or perform public health surveillance. Further, in Draft Principle 6, we do not believe that these types of surveillance research are rare. We suggest combining Principles 5 and 6, and removing the word “rare” for the following phrasing "In these instances, such as non-traditional settings like social media..." Collaboration Across Agencies A final consideration for this guidance is coordination across federal agencies. Research is increasingly funded through a variety of mechanisms, leaving institutions and researchers to face significant challenges in tracking different agency requirements related to data sharing, preservation, de-identification and re-use. One example of this is the chart already created for Certificates of Confidentiality, which may require multiple certificates to be requested for projects which have funding from a variety of agencies. We suggest that NIH coordinate federal agencies, particularly those involved with public health, in a collaborative effort to ensure requirements are consistent and do not put researchers in a position of conflict or replicating effort. Education As these policies and guidance come into effect, additional training and guidance as well as funding for creation of training and guidance is needed for researchers, students, institutional review board members, public and private partners, and potential research participants. We recommend further leadership by the National Networks of the Libraries of Medicine Data Services Office to identify and create the preliminary and ongoing resources needed. We additionally recommend the coordination of a regular open forum to document and identify ongoing challenges related to data security and privacy.

Email: pearsong@uic.edu

ID: 2066

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Jonathan Petters

Name of Organization: Joint organization RDAP and DCN

Type of Organization: Nonprofit Research Organization

Role: Scientific researcher

Comments:

In the attachment

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/ffhMUkmYrC.pdf

ID: 2067

Submit date: 6/27/22

I am responding to this RFI: On behalf of myself

Name: Challace Pahlevan-Ibrekic

Name of Organization: Northwell Health

Type of Organization: Health Care Delivery Organization

Role: Bioethicist

Comments:

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/KhYNMOoOEm.pdf

Description: Comment on DMS Policy

Email: cpahlevanibr@northwell.edu

ID: 2068

Submit date: 6/27/22

I am responding to this RFI: On behalf of myself

Name: David R Curry

Name of Organization: GFREI

Type of Organization: Nonprofit Research Organization

Role: Institutional official

Comments:

All sections ...see pdf attachment

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/AtRgWOFwtB.pdf

Description: GFREI Public Comment - NOT-OD-22-131

Email: david.r.curry@ge2p2global.org

ID: 2070

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Mary Jo Hoeksema

Name of Organization: the Population Association of America and Association of Population Centers

Type of Organization: Nonprofit Research Organization

Role: Scientific researcher

Comments:

Attached.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/MHQEKYPCEO.pdf

ID: 2071

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Margaret Levenstein

Name of Organization: ICPSR at the University of Michigan

Type of Organization: University

Role: Scientific researcher

Comments:

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/UTCZdhaRad.pdf

Description: ICPSR response to NIHRequest for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data Management and Sharing

Email: maggiel@umich.edu

ID: 2072

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Adam Rak

Name of Organization: Trustwave Government Solutions

Type of Organization: Other

Type of Organization-Other: Commerical Entity

Role: Member of the Public

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/wrnbByaqAs.pdf

Email: arak@trustwavegovt.com

ID: 2073

Submit date: 6/27/22

I am responding to this RFI: On behalf of an organization

Name: Anurupa Dev

Name of Organization: Association of American Medical Colleges

Type of Organization: Nonprofit Research Organization

Role: Scientific researcher

Comments:

Attached

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/WlfgGtVHDb.pdf

ID: 2074

Submit date: 6/27/2022

I am responding to this RFI: On behalf of an organization

Name: Duke University Offices: DOSI, DOCR, DCRI, RDI, Duke Libraries, Duke Medical Center Libraries

Name of Organization: Duke University

Type of Organization: University

Role: Institutional official

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/WkrIHMIVlJ.docx

Description: Duke University NIH DMS NOT-OD-22-131 RFC Response

Email: raul.doyle@duke.edu

ID: 2076

Submit date: 6/27/2022

I am responding to this RFI: On behalf of an organization

Name: Deborah Motton

Name of Organization: University of California, Office of the President

Type of Organization: University

Role: Institutional official

Comments:

In the attached comment letter, UC offers feedback on each of the aspects of the framework. In addition, we encourage NIH to continue to develop and support field-specific data repositories that would assist researchers in their data management and sharing efforts.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/fvRdBwZNuE.pdf

Description: UC Comment Letter

Email: agnes.balla@ucop.edu

ID: 2077

Submit date: 6/27/2022

I am responding to this RFI: On behalf of an organization

Name: Mandana Veiseh

Name of Organization: Polybiomics, Inc. / Lawrence Berkeley National Laboratory

Type of Organization: Other

Type of Organization-Other: Life Sciences/Biotech Startup

Role: Scientific researcher

Comments:

DRAFT Operational Principles for Protecting Participant Privacy When Sharing Scientific Data 2. Researchers and institutions should proactively assess appropriate protections for sharing scientific data from participants, including determining whether sharing should be restricted through controlled access,[4] regardless of whether the data meet technical and/or legal definitions of “de-identified” and can legally be shared without additional protections (e.g., the research does not meet the definition of “human subjects research” under the Common Rule). 3. Investigators and institutions should develop robust consent processes that prioritize clarity regarding future sharing and use of scientific data, including limitations on future use, and general aspects regarding how data will be managed (see Informed Consent for Secondary Research with Data and Biospecimens: Points to Consider and Sample Language for Future Use and/or Sharing).[5] Importantly, when a study offers the possibility of a direct benefit for research participants, the DMS Policy does not require sharing of data in order to participate. Comment: Modernizing data sharing policy and infrastructure enable a better foundation for bringing scientific data source closer to the users. And flexible evolution of this process by “involving investigators and institutions in development” provides many benefits. That includes reducing the delay that would result otherwise from assessing and implementing an overarching NIH policy. However, the exponential growth of the volume of medical and healthcare data that's being generated worldwide, from wearable devices (that monitor physiological to behavioral variables) and devices that monitor chronic conditions, demands constant evolution of powerful analytical approaches that might be unique to each institution. Also, because the word “appropriate” is subject to interpretation by each institute and thus can cause discrepancy in execution and outcome, the policy needs to specify the range of flexibility of operation within each institution, while clearly stating a governing guideline about all aspects of data privacy and data protection. This way, the investigators customize and utilize best practices within each institute and for each data type, effectively.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/zUKUnPoIrN.pdf

Description:

Email: mandanav@polybiomics.com

ID: 2078

Submit date: 6/27/2022

I am responding to this RFI: On behalf of an organization

Name: Luke Rasmussen

Name of Organization: NIAID/DMID Systems Biology Data Dissemination Working Group

Type of Organization: Other

Type of Organization-Other: Working group as part of NIH-funded research centers

Role: Scientific researcher

Comments:

Please see our comments in the attached document (Response to RFI NOT-OD-22-131.pdf).

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/ORypEqhCaB.pdf

Description: Response to: NIH Guide Notice NOT-OD-22-131

Email: luke.rasmussen@northwestern.edu

ID: 2079

Submit date: 6/27/2022

I am responding to this RFI: On behalf of an organization

Name: Dessi Kirilova

Name of Organization: Qualitative Data Repository

Type of Organization: Nonprofit Research Organization

Type of Organization-Other:

Role: Scientific researcher

Comments:

Please see attached response, which contains both general and specific comments.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/IYpehoQPTn.pdf

Description: 2022 QDR comments to NIH RFI on draft supplemental info

Email: Dessi.Kirilova@syr.edu

ID: 2080

Submit date: 6/29/2022

I am responding to this RFI: On behalf of myself

Name: Maya Sabatello, Daphne Oluwaseun Martschenko, Kyle Brothers

Name of Organization: Columbia University, Stanford Medicine, University of Louisville

Type of Organization: University

Role: Bioethicist

Comments:

The DRAFT supplemental information to the NIH Policy for Data Management and Sharing (herein Guidelines) calls for strengthening data sharing practices along with protection of research participants privacy and confidentiality. To further the goals of responsible data sharing, the Guidelines should: 1) assess the appropriateness of data sharing beyond individual participants and consent forms to include informal agreements with communities; 2) consider potential group harm from data sharing; 3) call attention to heighted protections for data collected through community-engaged research, especially marginalized communities; and 4) require researchers to make arrangements for benefit sharing. In addition, there is a need to ensure that researchers and reviewers of data sharing plans are knowledgeable about and considerate of issues of diversity, equity, and inclusion. More specific comments are below. 1. Focus on individual consent forms for delineation of data sharing practices is problematic. Studies consistently show that research participants have limited understanding of consent forms, and that they often sign the ICF with little attention to details. Consent forms also do not commonly include explanations of all types of sociopolitical harms that may arise from research—not only to individual research participants but also to entire communities such as stigmatization of certain populations. Moreover, the legalistic nature of consent forms does not fit well with relationships of trust that are critical in community engaged research. In community engaged research, many informal agreements exist and develop throughout the course of the study, especially as researchers learn more about the needs, values, and preferences of community members and partners; these new insights may vary significantly from what was originally envision in the consent form. The Guidelines should require researchers to consider the impacts of data sharing on communities at large. In the context of community-engaged research, the Guidelines should also allow flexibility for researchers and community members to revisit data sharing provisions that were included in the consent form to ensure that are in line with new insights as emerged during the study. 2. Institutional oversight for data sharing must consider risks for group harm and responsibility of secondary users. Institutions and IRBs that provide oversight on research studies have focused primarily on the protection of individual research participants, and the DRAFT Guideline only focus on delineating responsibilities regarding privacy and confidentiality. However, there is a need to ensure that data sharing agreements include clear consideration of risks for group harm and responsibility of secondary data users to ensure that their research does not exacerbate misperceptions regarding certain populations. For example, research to "identify" genetic influences on educational attainment helped motivate the recent mass shooting of 10 African American individuals in Buffalo. The Guidelines should require institutions to make arrangements to ensure that potential group harms are considered in decisions about data sharing as

well as that secondary data users are committed to socially responsible research. 3. Data collected in community-engaged research should be included as a justifiable exception to sharing scientific data. The DMS policy expects researchers to consider whether access to scientific data should be controlled. It includes key rationales for such control: legal requirements, sensitive information, non de-identifiable data, and technological risks to participant privacy. However, the current DRAFT does not include an expectation to control access based on the risk that sharing of data collected in community-engaged research may be used in ways that undermine the values of community-engaged research and increase stigmatization of the very communities that were engaged in the study. For example, the portrayal of people with disabilities in medical and scientific journals is often negative (even offensive) and reflects misperceptions of passivity, lack of competency, and deficiency, regardless of the views of people with disabilities themselves (e.g., wheelchair bound, mental retardation, hearing loss). On the other hand, community-engaged research with people with disabilities is grounded in understandings about the use of respectful language and practices that empower community members and aim to eliminate inequities in research and society. Allowing secondary data users to access data that were collected in community-engaged research without consideration of community preferences and values undermines the trust relationship that enabled the data collection and the expected outcomes for communities. The Guidelines should require that secondary data users are aware of community preferences when accessing data that were collected through community-engaged research and that they are committed to follow these preferences, including linguistic preferences, to avoid perpetuation of societal structures of subordination. 4. Arrangements for benefit sharing must go hand in hand with responsible data sharing. The reasons for data sharing are grounded in perceptions of such data as public good and intention for large datasets to improve the health outcomes of individuals and populations. However, data sharing without a co-commitment to benefit sharing is unlikely to yield such results. Benefit sharing can take many and different forms but should be included as part of the conversation. Moreover, data sharing without benefit sharing is further problematic when secondary users may be key players in the creation, and exacerbation, of health disparities. For example, blind/low vision individuals expressed high interest in participating in precision medicine research but reluctance to share their data with pharmaceutical companies that commonly fail to provide accessible drug labels and medical devices. At a minimum, the Guidelines should require secondary data users to commit to some form of benefit sharing, and that they follow practices and legal requirements, per the Americans with Disabilities Act, that are inclusive and accessible to diversely abled populations. 5. Researchers and reviewers of data sharing plans must have cultural humility and competency. Cultural competency has gained attention in the past few decades, but it has largely focused on “established communities”, as defined by majoritarian groups. As a result, less established or even unrecognized communities such as “Asians” and people with disabilities are rarely included in efforts to educate researchers about the values, preferences, and cultures of these communities. To ensure that decisions about data sharing or exemptions are informed, it is critical that researchers and reviewers of data sharing have humility and cultural competency to consider currently unrecognized marginalized populations.

Uploaded File: https://osp.od.nih.gov/wp-content/uploads/rfi2022_dms_sup/uploads/FShGxEIIhu.docx