

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

Slido Discussion

Changing the Culture of Data Management and Sharing - A Public Workshop

Day One: April 28, 2021

(From Lawal Abisinuola UMB Nigeria): What is the importance of data monitoring in public health and how can we ensure data security?

(From Brianna Lindsay, Institute of Human Virology University of Maryland): Quality data sharing requires resources (time and personnel), how can we ensure that we acknowledge the requirement of these additional resources and that there will be resources available that extend beyond the life of the grant?

Responses:

(From Alex Uzdavines): This is particularly important for trainees. Both to make it easier for folks to do secondary data analyses and current trainees who are more interested in research infrastructure than particular content areas

(From Brianna Lindsay, Institute of Human Virology, University of Maryland): Likewise, who will pay for the resources required to submit a data management and sharing plan in a grant application?

(Anonymous): At PCORI (funder), we provide our awardees with additional funds to support their time/effort to work with our designated repository - ICPSR - to curate the data to a high standard.

(From Brianna Lindsay, Institute of Human Virology, University of Maryland): ICPSR is great!!

(From Tobin Magle - UW-Madison): How will you encourage universities to incentivize data sharing by valuing them for things like promotion and tenure?

(From Joe Gay PhD, LICDC Ohio University, Ohio Alliance for Innovation): How many participants today?

Response:

(Anonymous): We have over 1200 people participating today.

(Anonymous): Would NIH consider supplementing the \$250k of modular budgets, to be used exclusively for data management? Right now, funds for data management "eat into" funds that would have been available for research.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): How will the NIH data sharing requirements align with existing and expanding laws related to data protection such as the European Union General Data Protection Regulation?

(From Brooks Hanson): DMP's are another section of grants. Why not improve the justification up front with "...explaining how the outcomes of the research, including data, software, materials, will advance science and be shared using leading practices"

Response:

(Anonymous): This is a great point!

(From Lynda Emel - Fred Hutch-HPTN): How do you envision the policy applying to NIH funded clinical trial networks? One Data Management and Sharing Plan per network? Or one per clinical trial? How will the effective date be applied? Newly funded trials or also ongoing trials?

(From Megan O'Boyle): Will researchers ever consider the potential "life of the data" beyond their current needs? As a rare disease advocate I would like to see Informed Consents to be broader to allow for data to be shared for future use and analysis by others

(Anonymous): Open Science is important and all but data preparation for public sharing (de-identifying, codebook, etc.) is extremely effortful and time consuming. How can universities provide services to facilitate this with tight budget everywhere?

(Anonymous): Many researchers submit proposals to a number of different federal agencies. Is there a plan to coordinate data management/sharing policy across federal agencies, instead of each one having different requirements and documents?

Response:

(Anonymous): I concur with this question, we do need to start thinking about standardizing our requirements, tools and templated documents

(Anonymous): Is the workshop being recorded and will we have access to the recording after the workshop?

Responses:

(Moderator): Yes, the webcast is being recorded and will be available on the website about 1-2 weeks after the event.

(From Alex Uzdavines): Ya, it'll be posted publicly in a few weeks

(From Adam Craig, Drexel U): Are there similar considerations for incentivizing sharing of analytical or (meta)data infrastructure software as for sharing of data?

(Anonymous): How do you encourage participation/interest in the data sets?

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(From Jeff Wickliffe): What active role can NIH/NLM play in helping researchers train international research partners so those data and specimens can also be more broadly available to all researchers and research participants?

(From Nicholas Horton): Lance Waller (Emory University, co-chair of the NASEM Committee on Applied and Theoretical Statistics) has a useful paper on the topic of data contributions: <https://peerj.com/preprints/3204/> (also published in TAS)

(From Curtis Walker): How can I learn more about NLM's ability to track and administer credit to authors for their contributions? I'm interested in learning more about what the speaker referenced.

(Anonymous): Has there been discussions about assisting small institutions that receive less than \$2-5M NIH funding/year and will most likely not have IP revenue?

(Anonymous): The NIH's help with addressing hurdles related to HIPAA, Common rule, and FDA regs would be appreciated.

(Anonymous): Other than data, does NLM have any strategic plans for sharing methods and supporting in silico models?

(Anonymous): I am thankful that one of my previous studies was not funded by NIH because it generated a very large data set that I published from for 2 decades. How do we retain the rights to use it 1st after we did all the work of collecting the data?

Response:

(Anonymous): Your approach means that other scientists could not use the data to produce results in less than 2 decades...i.e., you put your personal career priorities over the need for science to advance quickly!

(Anonymous): What is the NIH's view on the emerging concerns of indigenous groups around the world with regard to data sharing (i.e. the CARE principles)?

Response:

(From Sabrina McCutchan): We are very interested in CARE at DGHI since we work all around the world - would love to connect with others who are trying to implement CARE and similar frameworks more widely at their institutions.

(Anonymous): How does NIH plan to measure whether shared data is in fact being used?

(From Chris Stewart): I am wondering if healthcare delivery system leaders and legal personnel have been included as stakeholders in this discussion. Obviously patient privacy is a priority for all of us, but healthcare operations have very different goals.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

Response:

(Anonymous): Session V (tomorrow) will explore some of the legal issues.

(Anonymous): Do you plan to ask researchers what they would need/want to see in data sharing to mean they would want to re-use data? And what they want to use it for?

(Anonymous): Trials and Tribulations— 11 Reasons Why We Need to Promote Clinical Trials Data Sharing

<https://jamanetwork.com/journals/jamanetworkopen/article-abstract/2775663>

(From Marina Soares e Silva, Product Manager RDM, Elsevier): Atul, failed experiments remain something that many researchers see the value for. How can this data be shared from the start of the data life cycle in your view and what do you think this will mean for the metadata schema wrapping it?

(Anonymous): How do we mitigate researcher concerns about their data being reused, and the original researcher being scooped? We see researchers opting to share data with non-commercial use clauses - can this be considered truly open?

Response:

(From Sabrina McCutchan): Embargoes are the typical method re: scooping. Perhaps others know of alternate methods.

(From Natalie Meyers): Logistics question - when a speaker says "I'll share that link in the chat" where's the link appear to attendees/can it be shared back? Like the paper Atul just mentioned?

Responses:

(From Moderator): Paper link is in the chat

(From Moderator): To clarify, it's the link called "Trials and Tribulations"

(Anonymous): How does NIH plan to measure compliance to data sharing? If a researcher shared a truncated data with no documentation, what use is data sharing in this context? How can we avoid data dumping and encourage data sharing?

(Anonymous): Who should take on the responsibility of checking or validating that data has been shared in a reusable way? Or should we leave it the individual researcher to share their data in whatever way they think best?

(Anonymous): What aspects of the EU Data Spaces and proposed Data Governance Act should be modeled in the US context, to facilitate global research data exchange interoperability?

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Broad data sharing is a concept that researchers did not include in consents when looking historically. So how do we address that consent did NOT include for this type of sharing proposed today?

(Anonymous): Why is there a focus on data sharing statements rather than data citation? Data citations are more trackable than statements, and citations can be created for nonshareable research objects.

Response:

(Anonymous): Related, data doesn't have to be made public in order to mint a DOI.

(From Lynda Emel - Fred Hutch-HPTN): ImmPort is a great data repository, I would love to see something similar for HIV prevention clinical trials sponsored by NIAID.

(From Jeff Wickliffe): A more general question for the NIH and NASEM. What are the plans for harmonizing data management and sharing plans/platforms across federal agencies that sponsor research-NSF, EPA, NASA (e.g. remote sensing of air pollution) and others?

Responses

(Anonymous): I can't speak on behalf of any specific agency, but interoperability is important. The Global Earth Observatory is looking at the adoption of DSPs & DMPs.
Ethan McMahon, US EPA

(From Jen Wagner): What are the near-term and long-term plans for funding STS, anthropological, and human-centered design research to inform and understand the datafied cultures you are trying to promote and assess success in a democratized, inclusive way?

(Anonymous): Will NIH audit the data management plans stated in funded grants to ensure that the plans are followed and working as expected?

(Anonymous): Given the importance of negative results and their accompanying data, how can we encourage and reward the publication and sharing of these studies?

(Anonymous): Federation has issues. What happens when part of a query depends on a federated source that is offline - or responds very slowly?

(Anonymous): *Opening clinical trial data: are the voluntary data-sharing portals enough?*
<https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-015-0525-y>

(From Megan Potterbusch): For David Haussler, how would you suggest determining when a federated approach is best? Are there still circumstances you think that a shared Data Commons would be appropriate?

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Here is an overview focused on the Yale Open Data Access project:

<https://www.nature.com/articles/sdata2018268>

(From Geoff Lomax, CIRM): Recognition & Reward: How about monetizing data sharing history? Your sharing track record counts in the context of peer review and funding decisions. Assign a transactional value to the activity then institutions will support too.

(From Maria Teresa Napoli): the speaker is saying: "you can avoid duplicating data if data is already in repository": but has it been verified when deposited?

Response:

(Anonymous): Good question - for domain repositories with data curators or software that verifies data before it can be deposited, Q&A has been done; in addition, researchers would know where to look for the data -

(Anonymous): What is the ideal timeframe by which large amounts of data can be cleaned up/qc'd for public data sharing. Right now its 3 months after generation of data

(From Samuel): I have two questions about OpenNEURO's auto-curation feature: (1) can you elaborate on how that works (2) where can I find out more information about the auto-curation feature?

(From Carly, CZI): Do Elsevier journals allow data to be cited in the paper's references?

Response:

(Anonymous): Yes of course!

(From Megan Potterbusch): How often are you finding ORCID info for researchers associated with data metadata?

(Anonymous): Do you worry about combining datasets and the increased probability of making the data sets identifiable (which goes against CoC obligations)?

(Anonymous): Can the scientific community adopt metrics of how well an article cites its sources and data?

(Anonymous): Hello there, everybody! Will this workshop be videotaped and sent to participants, or will a connection be sent to participants so they can watch the portion of the workshop they skipped due to the time difference?

Response:

(Anonymous): Hello - yes, the workshop is being recorded and will be available for viewing on this website in several weeks.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Interesting to note that NIH has decided that program staff, rather than peer reviewers, will evaluate the suitability of DMPs. This reduces reviewer burden, but lends a worry about extent of subject area knowledge.

(Anonymous): How are FAIR different from prior collections of principles?

(From Lauren Maxwell): Can you talk about how to make DMPs machine readable, tools that exist for that?

Response:

(Anonymous): The DMPTool generates machine readable DMPs.

(Anonymous): Is there a format Bob suggests for using to make the DSP machine readable?

(Anonymous): Who is looking at principles beyond FAIR, such as TRUST & CARE, that address data ownership, etc.? The Group on Earth Observations promotes the implementation of these principles. Please contact me if you have ideas. Ethan McMahon, US EPA

(From Sabrina McCutchan): When papers are published using identifiable data that gets de-identified before it's deposited in a repository, how should this de-iding be best documented to facilitate reproducibility? Especially as whole variables may be omitted.

(Anonymous): Would including a track record of robust data sharing as part of scorable grant award criteria (points awarded for what's been done in regard to sharing data, not what is promised) be a strong incentive to write and comply with DMPs/DSPs?

(From Carrie): DMPTool v3.1 now has DMP IDs. See <https://blog.dmptool.org/>

(From Lynda Emel – Fred Hutch HPTN): NIH also has program staff review Statistical Analysis Plans....

(Anonymous): What is the direct connection between these two topics: data citations and DMPs? One is a grant requirement & one is a publishing convention. This seems very disjointed & I don't understand how they are related. Please help me understand.

(Anonymous): When we consider to evolve from machine-readable to machine-actionable DMPs, is it implicit the support of a workflow-based, dynamic DMP?

(Anonymous): There are tools like DMPTool.org that offer support for making machine actionable DMPs. Seems like that is a missing piece of this discussion.

(From Carrie): From DMPTool: ...moving DMPs from static text documents into structured, interoperable data that is able to be fed across stakeholders, linking metadata, repositories, and institutions, and allowing for notifications, verification, ...

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Bob: Accepting that there is concern about making DMPs/DSPs public in advance, is there an benefit/risk to making them public after the project has completed?

(Anonymous): RDMLA is financially sponsored by Elsevier.

(From Nina Exner): Many health science & academic libraries are under staffing cuts. Cuts are making it hard to provide data preservation, curation, & metadata services very robustly. Will PIs really budget time/effort to robust data librarian support?

(Anonymous): How do librarians ensure that publishers support data citations? They are involved in licensing content. Can they pressure publishers to support through those negotiations?

Responses:

(Anonymous): Yes, please! Raise open indexing of data citations with Crossref as an important priority for the library and something the publishers need to do.

(Anonymous): (Sorry it says anonymous -- this is Daniella Lowenberg who responded above)

(John P. Doyle, NIH Library): And increase indexing of repo datasets by CrossRef, Google Dataset Search, etc., to help surface the data not associated with preprint and/or published article.

(Anonymous): Does your training program work with The Carpentries? The wider library would benefit from working with RDMLA through Library Carpentry, and your content could benefit from informing researchers through Data Carpentry & Software Carpentry

(Anonymous): The Data Curation Network already does that for data curation:
<https://datacurationnetwork.org/>

(Anonymous): Many IRBs are resisting the use of GUID or any unique ID for data sharing , what do think about that?

(From Biru Zhou): I would like to know more about the technical difficulties for institutions who want to make DMPs interoperable into university systems (e.g., ethics applications, grant management, data deposits, etc.), particularly for admin purposes.

Response:

(Anonymous): There are RDA working groups that have defined a Common Standard for DMPs so that different systems can exchange the info in a DMP. It is the backbone of the features systems like DMPtool.org use.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(From Darren Kinsman): A key part to capturing data and making it shareable is a solution for researchers that provides ease of use - what technical solutions are used on the front end for researchers, and on the back end for a variety of data structures?

(From Megan Potterbusch): Implementing interoperable DMPs at a university or institutional level has primarily been a challenge in my institution through lack of admin awareness and barriers to selecting tools that facilitate interoperability or that prioritize DMP.

Response:

(Anonymous): The work on interoperability for DMPs has been a major focus of RDA and the RDA Common Standard Working group. They have a standard metadata application for DMPs - <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

(Anonymous): That has been done. There is a JSON format for interoperability. RDA working group did that already

Responses:

(From Biru Zhou): Could you please expand on which working group from RDA specifically?

(Anonymous): <https://www.rd-alliance.org/group/data-type-registries-wg-2/wiki/work-json-implementation-enriched-data-type-model>

(Anonymous): The work on interoperability for DMPs has been a major focus of RDA and the RDA Common Standard Working group. They have a standard metadata application for DMPs: <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard>

(Anonymous): Please mention DMPtool--I'm not sure why this hasn't been at the very least name-checked...

(Anonymous): The DMPTool follows the RDA Common Standard for DMPs and has an API to exchange this information with external services

(Anonymous): Societies need to become more involved. ACS and AGU have done a good job. Others in our field are way behind and even resistant to session proposals in this area.

(Anonymous): Can researchers be encouraged to share by explicitly prompting them to include evidence of data sharing within their biosketch?

Response:

(Anonymous): I agree. NIH could add this to their biosketch format!

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): I agree. This is a particular interest of mine as a principal investigator and I am familiar of A WAY to do it in my team. However, when I went to my cancer center, they actually do not have help available so each group is making it up.

(From Robert R. Downs, CIESIN, Columbia University): How can we help promotion and selection committees recognize the value of data sharing so that those who create data will see the value of data sharing for their careers?

(Anonymous): What tools can help with adding metadata (preferably based on discipline-specific standards) at the front end?

(From Nicholas Horton): The TIER (Teaching Integrity in Empirical Research) symposium (co-sponsored with Sheffield Methods Institute) has some relevant discussions:
<https://www.projecttier.org/fellowships-and-workshops/2021-spring-symposium/>

(From Alex Uzdavines): Earlier the idea of paying people to do data management runs into direct/indirect cost problems. Would ironing these out allow investigators to hire data managers work better than adding yet another task that investigators need to do?

Response:

(From Alex Uzdavines): Like.... if you make it feasible for the glut of PhDs getting produced to be hired (at a full professional salary) to do this work, it'll get done

(From Nicholas Horton): Nudge points and best practices are also discussed in Broman and Woo's "Data Organization in Spreadsheets" paper:
<https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989>

(Anonymous): Your comment on cataloguing is interesting. How are you approaching a collection of datasets that inform a study versus the individual data sets that make up the study, but each have the potential for reuse?

(From Eleni Wits): How does data management plans tie in with the ethics approval processes?

(Anonymous): How do the panelists suggest approaching situations where authors are reluctant to (or even refuse to) share their data? Even when asked?

(Anonymous): NIMH has developed models and clear sharing guidelines and Databases (combined NDAR and NDCT.. under NDA)

(Anonymous): The Chemical Information section of ACS hosted two impressive sessions: one on FAIR Data and another on Cultivating Good Data Practices Among Chemists. A few presentations highlighted librarian/faculty collaborations.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(From Philip Tarrant, Arizona State University): We have an RDM class that we teach once a year to introduce the topic to our grad students

(Anonymous): F, K and T32 training mechanisms can require training and address demonstration of data sharing/reuse.

(From Megan Potterbusch): Do you have any recommendations for identifying researchers and PIs who are not well-versed in good data management practices? Sometimes targeting outreach around training is a challenge

(From Tarek Sobeih, Nathan Kline Institute): We have a team dedicated to data management and sharing but funding is always an issue. Training grad students / investigators is challenging

(Anonymous): Administrative cuts to save \$ and fund more grants are hindering data sharing and curation efforts. Would have to fund fewer grants. What about leaning on Institutions to support data sharing efforts with Indirect Costs.

(Anonymous): SfN has an NINDS grant program that includes resources and conversations about how to institute a culture of rigor (including data management and sharing) in the community: <https://neuroonline.sfn.org/FRN>

(Anonymous): AAU and APLU in US just released a guide for institutions to accelerate public access to research data. Guide and registration links for conversations are at: <https://www.aau.edu/accelerating-public-access-research-data>

(Anonymous): Conversations about their recommendations are scheduled for May to learn more - one of the conversations focuses on harder problems, like P&T.

(Anonymous): The HHS Office of Research Integrity announced a new funding opportunity: Considerations, Options, and Resources for Data Management in PHS Funded Research (IR-ORI-21-001). <https://ori.hhs.gov/blog/ori-announces-two-funding-opportunities>

(From A Central Resource): Policy evaluation research is becoming much more complex, e.g. large scale govt. PII databases. Such work requires a board range of expertise. How best can researchers acquire the necessary complementary resources?

(Anonymous): But even if university provides these services - faculty need to pay for it internally so that does not solve the funding issue

Response

(From Alex Uzdavines): plus if you're working in a soft-money/government research center, this issue is even more glaring

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Academic institutions starting to think about this at the institution-level: Guide to Accelerate Public Access to Research (<https://www.aplu.org/projects-and-initiatives/research-science-and-technology/public-access/>)

(Anonymous): FOIA requires researchers with Fed grants to share

(From A Central Resource): Could we get a link to Manchester's data management service cited by Professor Goble

Response:

(Anonymous): <https://society-rse.org/>

(Anonymous): FOIA share data for 3 years after closeout of project

(Anonymous): I propose that there should be a requirement that the data sharing plan has to be specific, with milestones, and PI is required to specify in their progress report of how they meet the milestones. Type 5 award will be issued accordingly

(From Carrie): There used to be a supplemental award for Informationist Services in NIH-funded Research Projects. I.e., adding a data librarian to your research team.
<https://www.nlm.nih.gov/ep/InfoSplmnts.html>

(Anonymous): My favorite "data have to be born FAIR"

(Anonymous): "Cross sector collaborations could assist with scientific trust and scientific literacy. What would this look like? How can this change in culture help to translate big data resources into tangible interventions?"

Day Two: April 29, 2021

(From Bella Gutnik, Atrium Health): Are you going to send slides from the conference via email? When should we expect it? Thank you!

Response

(Anonymous): The slides will be available on this website in the next 1-2 weeks. Thanks!

(Anonymous): Will yesterday's recording be made available?

(Anonymous): What if the researcher says Yes to all the questions: that doesn't make the data available in a repository...

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

Response:

(Anonymous): Good catch. The questions are meant to start a conversation about data management (which does not, necessarily, result in sharing). If they answer yes to all, the follow-up is "Great, let's talk about your options for sharing". -John Borghi

(Anonymous): How important is research experience in a data steward professional? It seems like that lens is helpful in posing the kinds of questions John B poses.

(Anonymous): Yesterday it was mentioned that some researchers were embarrassed (data sharing) or didn't want their data and findings scooped? any comments?

(From Lars Vilhuber): "Data Management" is often taught - see Data-science courses. But often, the data that is managed is how to "acquire" data (web scraping, etc.). The term is overlaid, and that may lead to confusion/lack of training.

Responses:

(Anonymous): Completely agree! When I talk with researchers, I quickly pivot to "data organization", "documentation", etc. It would be fantastic if we could find a compelling alternative term.

(From Natalie Meyers): sometimes effective alt terminology is Research Data Curation, Research Data Stewardship, Research Data Preparation for Sharing and Citation . . .

(Anonymous): What effect has the incorporation of ELNs into the workflow on good data management practices?

Response:

(Anonymous): It can be very helpful to have tools like ELNs to make adoption of best practices easier, but I'm fond of saying that good data management ultimately comes down to behaviors rather than tools. -John Borghi

(From Nicholas Horton, Amherst College): The data management practices highlighted in today's presentation dovetail well with the NASEM (2018) "Data Science for Undergraduates" report's components of data acumen. We need to be teaching these foundations early and often.

Response:

(Anonymous): Absolutely agree. I never learned data management in a formal setting and I am a strong advocate in getting it integrated into research methods-related curriculum whenever possible. -John Borghi

(Anonymous): How can journal staff support and promote outreach? Are there free resources that we can share with our authors?

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Doesn't public data also help accelerate science, e.g., vaccine development? Yesterday someone commented that they preferred to mine their data for 20 years, rather than letting others use the data to advance science probably more quickly!

(From Shari Leventhal, American Society of Nephrology): How can journal staff support and promote outreach? Are there free resources that we can share with our authors?

(Anonymous): Is there a good tool to help foster data literacy, or strategies to help foster engagement?

(Anonymous): The Census Bureau has advanced this research substantially since 2010. Their privacy team gives wonderful presentations if anyone is looking for more info...

https://www.census.gov/about/policies/privacy/statistical_safeguards.html

(From Lynda Emel, Fred Hutch HPTN): In what circumstances can birth date be included in a data set for sharing? We usually remove birthdate and replace with age at enrollment for clinical data sets sharing, but we have one investigator who insists birthdate must be included

Responses:

(From Lars Vilhuber): (a) share privately/ with controls (b) ask what birthdate is needed for (link to contextual data?), and do some custom (private) merging.

(From Lynda Emel, Fred Hutch HPTN): Thanks Lars - we now share privately with controls with investigators who have approvals from protocol team and can continue that. Also had issues with census tracts and plan to provide special data of interest related to census tract w/o

(From Lynda Emel, Fred Hutch, HPTN): w/o actual census tract number.

(From Lars Vilhuber): You may still want to worry about secondary re-identification - adding census-tract level information but stripping tract number does not prevent re-id of the tract...

(Anonymous): Will the presentations be available online after the workshop?

(Anonymous): What was the reference for the Pasquetto data sharing publication?

Response:

(Anonymous): Pasquetto et al (2019) Uses and reuses of scientific data: the data creators' advantage, Harvard Data Science Review, 1(2)

(Anonymous): When should researchers, or their data stewards, be interacting with their institutional privacy officers? As a CIPP/US, I recognize there is so much value they can add as privacy experts, yet they aren't represented here

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): Regarding costs - How are you thinking about striking the balance to make raw data available, such as image data that can be quite large, with analyzed data?

Response:

(From John P. Doyle, NIH Library): Great example - with diagnostic-type images, uncompressed, verifiable (e.g. checksum) masters will often be required - and storage and access costs are not trivial here.

(From Lynda Emel, Fred Hutch HPTN): Even completely de-identified data can be used to publish unethical or unfair or unlearned opinions about certain groups of people.

(Anonymous): What are the ethical considerations or how do you share data broadly when consent/authorization is not obtained from the participant (e.g. waiver of consent and/or authorization)?

(Anonymous): Given the concerns for Tribal data, what are the panelist's opinions on whether this sensitivity is appropriate to extend to all population data?

(Anonymous): How do Certificate of Confidentiality requirements coordinate with broad data sharing, especially if the data is identifiable?

(Anonymous): How is security handled in the enclave to make sure that the bring your own data or bring your own pipeline are appropriately addressed/assessed for security and appropriate data use?

(Anonymous): Excellent presentations - will the speakers' slides be available?

Responses:

(Anonymous): Yes, most of the presentations should be available on this webpage within two weeks.

(Anonymous): Thank you!

(From Christina Drummond): The NIST Privacy and Cybersecurity Frameworks (and the crosswalk between them) are really helpful. How could NIH leverage them and the developing NIST RDAF to increase researcher awareness of privacy and cybersecurity standards?

(From Nicholas Horton, Amherst College): Here's a link to the Dunlop et al paper "The Impact of HIPAA Authorization on Willingness to Participate in Clinical Research", Annals of Epi, 2007, <https://www.sciencedirect.com/science/article/abs/pii/S1047279707002098>

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

(Anonymous): How do we deal with data collected from Egyptian mummies without asking families of the pharaohs or Egyptians? Will the mummies be returned to Egypt? Will we stop using technology developed from information collected from these mummies?

(Anonymous): Sharing of the data can also be tracked metrics, if there are DOI that can be tied to the data.

(From Samuel): The Wellcome Trust has an early career researchers pack to inform researchers about data sharing best practices. Does the Bill and Melinda Gates foundation have a similar thing in place in order to promote good data sharing practices?

(From Natalie Meyers): Ashley, Your slides are rich with hyperlinks to resources and examples that are of great interest to me, and I imagine to all of today's attendees, could you share the slides in a format with us so we can visit some of those hyperlinks?

Response

(Anonymous): Yes, the slides will be available as PDFs on the website in approximately two weeks.

(Anonymous): Since we were 20 mins behind on agenda @ start of this talk, will we have the next break or skip it?

(From A Central Resource): Institutions often make significant funding into physical and lab infrastructure but much less to support data management, analysis and dissemination. Is this an institutional problem?

(Anonymous): Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. Harvard Data Science Review, 1(2).

<https://doi.org/10.1162/99608f92.fc14bf2d>

Response:

(From Rob Crystal-Ornelas): Thanks for sharing this link/article!

(Anonymous): +1 for funding RDM/data sharing training programs (comprehensive and disciplinary specific).

(From Natalie Meyers): Phil mentioned an AAU/APLU report that "points the way" This one?

<https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Public%20Access/AAU%20APLU%20Guide%20to%20Accelerate%20Public%20Access%20to%20Research%20Data.pdf>

Responses:

(From Natalie Meyers): Association of American Universities and Association of Public and Land-grant Universities (2021).

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

Guide to Accelerate Public Access to Research Data. Washington, DC.

DOI: <https://doi.org/10.31219/osf.io/tjybn>

(Anonymous): Yes, Presented on this site: <https://www.aau.edu/accelerating-public-access-research-data> (and "published" on the link you have in your reply).

(Anonymous): Outreach to institutions is continuing through a series of 1-hr conversations in May - register at <https://www.aau.edu/accelerating-public-access-research-data>

(Anonymous): Just mentioned: great book on "whose data," includes much on health equity issues. Won book awards in econ and finance: Perez, C. C. (2019). Invisible Women: Data Bias in a World Designed for Men (First Printing edition). Harry N. Abrams.

(Anonymous): Couldn't data sharing help promote equity by bringing transparency to the data (and methods and practices) that is being used in research?

(Anonymous): Could all the resources mentioned in the chat also be shared with the recordings please?

Response:

(Anonymous): We can clean up and share the resources from the chat in the next couple of weeks. Thanks!

(Anonymous): What is a good example from any organization in the world of good data management and/or stewardship?

(From Adam Craig, Drexel University): How do you expect institutions to provide educational initiatives on data sharing when the technologies and organizations covered may no longer be relevant by the time the curriculum gets written and approved?

(Anonymous): agree at the qualitative. In-depth conversations can get at a lot of nuanced perspectives.

(Anonymous): Can the panel speak to specific actions that researchers can take to engage PEER communities as full partners in research, with the objective of promoting equity?

(Anonymous): Participatory approaches to research (e.g., CBPR) provide broad guidance on community engagement but are not specifically focused on PEER communities or equity as an end goal.

(From Karen Wehner, HHS Office of Research Integrity): The HHS Office of Research Integrity has announced a new funding opportunity for research and development of practical approaches and resources/tools for data management.

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

<https://ori.hhs.gov/blog/ori-announces-two-funding-opportunities>

(Anonymous): Comment re: having NIH walk the talk. There is incredibly diverse willingness to submit DMPs and share data by the NIH intramural scientists. Most NIH discussions put emphasis on the extramural community.

Additional Resources Shared During the Workshop by Planning Committee Members and Speakers

The 27th annual Nucleic Acids Research database issue and molecular biology database collection (Daniel J Rigden, Xosé M Fernández):

<https://academic.oup.com/nar/article/48/D1/D1/5695332>

“Secondary Analyses” RFAs:

- Informatics Methodology and Secondary Analyses for Immunology Data in ImmPort <https://grants.nih.gov/grants/guide/pa-files/PAR-19-229.html>
- NEI Research Grant for Secondary Data Analysis <https://grants.nih.gov/grants/guide/pa-files/PAR-09-252.html>

WSJ article, *Some California Hospitals Refused Covid-19 Transfers for Financial Reasons, State Emails Show*: https://www.wsj.com/articles/some-california-hospitals-refused-covid-19-transfers-for-financial-reasons-state-emails-show-11603108814?st=r5xpnz4jh24vgk8&reflink=article_email_share

In Pursuit of PPE: <https://www.nejm.org/doi/full/10.1056/NEJMc2010025>

Michael Bramhall, MSc, Oscar Flórez-Vargas, MSc, Robert Stevens, PhD, Andy Brass, PhD, Sheena Cruickshank, PhD, Quality of Methods Reporting in Animal Models of Colitis, *Inflammatory Bowel Diseases*, Volume 21, Issue 6, 1 June 2015, Pages 1248–1259, <https://academic.oup.com/ibdjournal/article/21/6/1248/4579338>

Global Biodata Coalition: <https://globalbiodata.org/>

The Value and Impact of the European Bioinformatics Institute (Jan 2016): <http://www.beagrie.com/EBI-impact-report.pdf>

Cost of Not Having FAIR Research Data:

http://publications.europa.eu/resource/cellar/d375368c-1a0a-11e9-8d04-01aa75ed71a1.0001.01/DOC_1

Williams, Heidi L. "Intellectual property rights and innovation: Evidence from the human genome." *Journal of Political Economy* 121.1 (2013): 1-27. <https://www.journals.uchicago.edu/doi/abs/10.1086/669706>

These questions and comments were shared online by participants during the live workshop, Changing the Culture of Data Management and Sharing, on April 28-29, 2021. The discussion is now closed.

Gold, E. Richard, The Fall of the Innovation Empire and its Possible Rise Through Open Science (March 17, 2021). Research Policy 50:5 104226 , Available at SSRN: <https://ssrn.com/abstract=3813139>

Data Governance Task Force: Final report and recommendations (2016): <https://Ucla.App.Box.Com/s/Zvtg5rcd9ojo2by2rtzoum4kv9mdulnr>

UCOP Privacy and Information Security Initiative (2013): <http://ucop.edu/privacy-initiative/>

Hilgartner, S., & Brandt-Rauf, S. I. (1994). Data Access, Ownership, and Control: Toward Empirical Studies of Access Practices. Science Communication, 15(4), 355–372. <https://doi.org/10.1177/107554709401500401>

From Rick Gilmore:

<https://github.com/psu-psychology/psy-525-reproducible-research-2020>
<https://github.com/psu-psychology/r-bootcamp-2019>
<https://www.youtube.com/watch?v=N2zK3sAtr-4>