

# **“PROTECTING PEOPLE, SHARING DATA”**

## **DATA SHARING AND HUMAN PARTICIPANT CHALLENGES IN THE SOCIAL SCIENCES**

May 21, 2018  
LUBIN HOUSE, NEW YORK CITY

### **BACKGROUND / FRAMING NOTES**

Social scientists are increasingly expected to make their research transparent, and to maximize the value of data by making them available to others, whether in pursuit of research transparency or for secondary analysis. Important questions remain about how sharing data impacts social science’s ability to produce more credible and legitimate knowledge. Nonetheless, there is undeniable and significant momentum towards instantiating openness as the default scholarly position, with exceptions based on established grounds.

Chief among these grounds is the obligation to protect human research subjects. A foundational premise of social inquiry is that there are important questions that can only be answered by engaging with human subjects, and then analyzing the information such encounters offer. Social scientists have long recognized that this engagement involves risks for human subjects, and various arrangements have been made to assess and address these concerns, including promising subjects anonymity, and guaranteeing that the information they provide to a research team will not be shared more widely.

Institutional Research Board (IRB) are charged with helping social scientists conduct research in ways that will protect their human subjects from harm. In that role, IRBs sometimes encourage scholars to withhold – or even destroy – their research data in an effort to prevent even minimal risks of disclosure of private information. As interest in data sharing and research transparency has grown, IRBs and other institutions in the social science ecosystem – in particular funding organizations and academic journals – have engaged in rich discussions about how to continue protecting research participants while simultaneously making academic research more open.

As the momentum towards openness has grown, and as funders and publishers have begun to consider questions of data sharing more deeply, they have started to delineate systematic arrangements for the sharing of certain types of data and for making particular kinds of research more transparent. They are developing internal mechanisms, aligning infrastructure, and promulgating guidance for what data their grantees and authors need to provide, and when they need to provide it. These systematic arrangements are consistent, predictable and transparent. They allow researchers to plan and conduct research with shared expectations of what funders and journals in their respective fields are likely to require.

Yet to date few funders or journals have developed or deployed clear procedures or arrangements for sensitive data generated through interaction with human participants. While these stakeholders’ requirements and arrangements for data sharing invariably *list* and allow exceptions for sensitive data that might put human subjects at risk, their systematic processes for sharing research data currently lack procedures, for instance, for determining whether and how

they grant exceptions to their data sharing requirements based on data sensitivity. We hope to encourage and offer our support in the creation of such procedures.

To be clear, we are not suggesting that either funders or journal editors adopt a one-size-fits-all approach to working with sensitive human participants data. Likewise, we are not proposing that funders and journal editors undertake substantive review of grantees' or authors' arrangements to protect human subjects. While funders and journal editors are keenly interested in the tensions between protecting human subjects and sharing research data, determining when the former should limit the latter is neither their main interest nor core function. Funding organizations are primarily concerned with the substance of the research activities they support. They encourage or require data sharing because such sharing ensures that research is as rigorous and valid as possible and facilitates data reuse, thus increasing the return on their investment. Similarly, journal editors' first concern is the contribution a publication makes; they require data sharing and research transparency to help ensure scholarship is well-founded and can be understood and evaluated. Nonetheless, for both types of stakeholders, human subjects protections represent a universally recognized reason for limiting data sharing.

What we do propose is that funders and journal editors develop clearly described sets of procedures for how scholars can bring their concerns about sharing research data to these stakeholders' attention. At what stage of their funding application or manuscript submission will scholars be encouraged to do so? What documentation will be required? And so on.

We also believe it would be beneficial for stakeholders to acknowledge the utility of more variegated arrangements for sharing sensitive data. Rather than seeing sharing data in "all or nothing" terms, the goal should be making data "as open as possible, *as closed as necessary*."<sup>1</sup> Achieving that goal entails introducing more nuanced solutions that allow "some" data to be shared: scholars might share subsets of their data, de-identify data, use credentialing, or place access controls on the data. The idea, then, is to generate more differentiated answers to the questions of "what" data can be shared, "how" they can be shared, and "with whom" they can be shared.

This more nuanced thinking about differentiated arrangements will undoubtedly make each stakeholder's interactions with researchers (and their data) more complex. Yet the potential payoff is considerable. Counterintuitively, introducing additional arrangements for *restricting* data access will allow *more* data to be shared. Where the default choices are sharing all of one's data or sharing none, even the slightest risk leads a researcher to share none of her data. Where sharing *some* data (with protections) is a feasible option, sensitive data that would otherwise have been unavailable can be shared.

The table below illustrates the dual shifts just mentioned. Before conversations about data sharing and research transparency began to pick up again in the social sciences over the last decade, to the extent that funders' and journals' practices required data sharing, they likely fell into cell (A). They treated data sharing as an all-or-nothing proposition, and they dealt with exceptions on an ad hoc basis. Given the direction of current conversations, we see cell (D) as a

---

<sup>1</sup> ERAC 2016, 15 emphasis added; see also European Commission Directorate-General for Research & Innovation 2016, 6.

possible and preferable destination. Scholars, stakeholders, and the social science enterprise more generally would benefit from funders and publishers adding systematic procedures for responding to scholars’ concerns about sharing sensitive research data. The scholarly enterprise would likewise benefit if they expand the repertoire of acceptable solutions to the challenge of sharing sensitive research data to include a more nuanced set of restricted data sharing arrangements.

		Data sharing not required	Date sharing required but with exemptions for sensitive data, where choices are:	
			Dichotomous (all or nothing)	Nuanced (different degrees)
Way in which requests to except sensitive data from sharing are processed	Ad hoc	<i>Empty set</i>	(A) Previous status quo	(B)
	Systematically	<i>Empty set</i>	(C)	(D) Possible Future

Our goals for this workshop (as well as the other events in our broader initiative) are to encourage the various stakeholders in the social science ecosystem (e.g. scholars, ethics committees, funders, publishers, and data repositories) to:

- consider and discuss the practicalities and implications of this dual shift; and
- jointly construct a framework to help researchers navigate the tensions among learning from human subjects, protecting them, and sharing information generated through interaction with them.

Joint engagement on these issues is critical, because the tensions between sharing sensitive data and research openness arise at multiple points in the research cycle – as scholars garner IRB approval, apply for funding, conduct research, and publish. Our specific objectives are to: (1) increase stakeholders’ knowledge of the roles, responsibilities, and perspectives of the others, and (2) encourage each stakeholder to produce clear, coherent and meaningful guidance that is in harmony with the others’ directions.

The range of guidance that IRB’s typically provide through their prior decisions is a crucial factor in how other stakeholders will approach the dual shift of systematizing human subjects exceptions for data sharing and allowing more nuanced approaches to contingent data sharing. Some forward-looking IRBs have begun to consider how the imperatives of data sharing, research transparency, preserving confidentiality, and securing informed consent will all interact. For instance, Cornell University’s IRB’s recently revised the consent script language it offers to its social and behavioral researchers in very encouraging ways. In a dedicated “Data Sharing” section, the suggested wording is premised on the understanding that data will be made available in an appropriate form. Importantly, the wording directly invokes two critical tools for managing the risks that sharing sensitive qualitative data can pose: de-identification and controlled access.

More members of the IRB community are indicating a willingness to move in this direction. As one example, following a previous workshop in this series attended by representatives from IRBs at 18 U.S. research universities, a subset of the attendees co-drafted suggested guidance and

consent script language that could facilitate the development of such text by other IRBs. Ideally, funders and journal editors will take note of developments like these in both IRB and data management communities as they revise their policies and guidance concerning the sharing of sensitive data so that the information and advice that scholars receive as they traverse the research cycle is consistent and aligned.

Researchers also play a crucial role in recognizing and mitigating the risks that can be associated with sharing the sensitive data their work generates. On the one hand, under the current regulatory system of research, scholars generate a research application that an IRB reviews and approves. On the other hand, scholars are increasingly called on – by funding organizations, for instance – to develop data management plans (DMPs) that include a discussion of the security of the data they expect to collect and that consider attendant questions of ethics, privacy and legal responsibility. Stakeholders’ increasingly sophisticated – and harmonized – thinking about openness should be incorporated into the structure and review of scholars’ IRB research application and of their DMPs. Scholars should plan to use – and reference in the applications and plans – emerging technologies that allow them to optimize the relationship between protecting their human participants and sharing their research data.

In sum, social science has taken very promising first steps toward a new equilibrium in which data sharing—with all of its positive externalities—becomes more feasible. There is a new willingness on the part of multiple stakeholders within the social science community to engage with each other with the goals of generating new consensuses on the safe and appropriate sharing and publishing of sensitive research data, as well as developing clear, consistent and meaningful guidance for scholars to do so. Accordingly, now is a critical moment to begin to work out the requisite practices and procedures. Social science data repositories, including those gathered at this workshop – Dataverse (Institute for Quantitative Social Science, Harvard University), ICPSR (Institute for Social Research, University of Michigan), and the Qualitative Data Repository (Center for Qualitative and Multi-Method Inquiry, Syracuse University) – can serve as key resources for these efforts.

As the agenda below outlines, we take the draft guidance and informed consent language referred to above as the starting point for the workshop. After the introduction, sessions one, two, and three each consider specific data scenarios (based on how IRB documentation deems the data with regard to sensitivity and shareability). For each scenario we will consider a series of functional questions to which funders and editors will need sets of answers so they can systematically, fairly, and equitably deal with scholars’ concerns about sharing sensitive research data. How will funders and editors address the challenges that arise in the various scenarios, and how will they build systems to accommodate the solutions? While we will rely on workshop attendees from the IRB community to offer their expertise, guidance, and perspective throughout the workshop, in the fourth session will invite them specifically to react to the discussion so far. In a fifth session we will think together about the practical implications of the workshops’ themes and discussions for social science researchers.

**“PROTECTING PEOPLE, SHARING DATA”**  
**DATA SHARING AND HUMAN PARTICIPANT CHALLENGES**  
**IN THE SOCIAL SCIENCES**

May 21, 2018  
LUBIN HOUSE, NEW YORK CITY

**AGENDA**

**8:30-9:00 Light Breakfast**

**9:00-10:00 Motivation for and Scope of the Workshop**  
**QDR Presenter: Colin Elman**

This introductory session will cover the goals of the workshop: to consider how norms guiding the sharing of social science data are changing, to think through the challenges that sharing data presents for scholars who directly engage human participants (research subjects) in creating or collecting data, to consider the costs of *not* sharing sensitive data when it is possible to do so with minimal risk, and to think through how the various stakeholders of the social science community can re-envision these norms, meet these challenges, and mitigate these costs. The overall objective is to begin to build a framework for social science stakeholders to navigate the tension between protecting human participants and maximizing the value of data. We will also discuss how the workshop builds on previous meetings and the materials that they generated. Each workshop participant will be asked to give a two-minute self-introduction.

**10:00-10:15 Coffee break**

**Organizational note:** In the next three sessions we will discuss four ideal-type scenarios involving sensitive human subjects data, and consider how funders and publishers might generate systematic arrangements to accommodate the data in each. The first three scenarios involve circumstances reflected in the illustrative guidance language produced through discussions with senior IRB staff at the previous workshop. In the fourth scenario – one in which we believe researchers will often find themselves in the near term as practices change more quickly than do formal policies, there is a mismatch between what IRB documents suggest and what grantees/authors wish to do with regard to data sharing.

**10:15-11:00 Scenarios One and Two: “Unshareable” and “Unproblematic” Data**  
**QDR Facilitator: Dessislava Kirilova**

The first two scenarios are the most straightforward with regard to the status of the data. In the first, a scholar approaches a funder or an editor with IRB documentation indicating that her data cannot be shared with anyone under any circumstances without putting human participants at undue risk, and the scholar’s assessment of the situation accords with that evaluation. Perhaps the scholar was embedded with a rebel army and garnered crucial information about a conflict from those participants. In the second scenario, IRB documentation deems the data in question to be unproblematically shareable, and the scholar concurs; perhaps she interviewed a small set of elites who agreed and in fact insisted on giving the interview on the record and attributed. How will a funder with a data sharing requirement deal with each of these scenarios? How will they be accommodated in a journal’s workflow?

**11:00-11:15 Coffee break**

**11:15-12:00 Scenario Three: Contingent Data Sharing**  
**QDR Facilitator: Diana Kapiszewski**

In the third scenario, a scholar approaches a funder or a journal editor with IRB documentation indicating that some data are shareable if precautions are taken. Perhaps the funder or journal requires that data be deposited in a particular venue (e.g., their own dataverse), but it does not offer the functionality required by the contingency. How will stakeholders manage cases like these? Even when contingencies can be accommodated (because the stakeholder’s venue offers the functionality, or a third party institution is permissible), are all contingencies acceptable? For example, would the stakeholder agree to data access being conditioned on the scholar’s permission, allowing her to pick and choose which (potentially critical) users see the data? Who will decide the conditions under which the data can be shared? How will this scenario be dealt with by funders, and in a journal’s workflow?

**12:00-1:00 Lunch break**

**1:00-1:45 Scenario Four: Evolving Evaluation of Data Shareability**  
**QDR Facilitator: Dessislava Kirilova**

In this session we consider a series of scenarios in which a scholar wishes to act in a way that diverges from what her original IRB documentation stipulates with regard to data sharing. We anticipate that these types of challenges will arise quite frequently in the near term, as social science stakeholders’ expectations about responsible data sharing evolve, technologies facilitating such sharing emerge, and practices realign to match those expectations and capitalize on those technologies. In each scenario, a scholar – now having conducted her proposed research and perhaps with a new understanding of strategies for sharing sensitive data – evaluates the shareability of her data differently from how she did so when she developed her IRB materials

(either considering them to be more *or* less sensitive). In one scenario, the scholar anticipated the sensitivity of the data and indicated a corresponding sharing strategy to her IRB (and her human participants). In a second scenario, a scholar could not anticipate how sensitive the data would be and thus did not discuss data sharing in her IRB application (or with her human participants). In what directions, in what ways, and when can a researcher deal with different original IRB stipulations regarding data sharing and how will these changes be dealt with by funders and editors? When are amendments to an IRB application appropriate – or sufficient – and under what circumstances is re-contacting human participants recommended or necessary? How does each process interact with funders’ and journals’ procedures and workflows?

**1:45-2:30      Reflections from the IRB Community**  
**QDR Facilitator: Dessislava Kirilova**

While experts from the IRB community will inform all of our discussions, in this session we offer them the floor to share their reflections on and reactions to the conversation thus far. How do they see – and how will their community see – funders’ and editors’ preferences, and potential policies and practices, with regard to scholars sharing their sensitive research data? For instance, are they comfortable with funders or editors considering the content of IRB documentation for the purposes of exempting scholars from sharing data? If this piece of documentation that is usually meant to mediate between researchers and human participants is to be used for this purpose, will it need to change in any way? How can we ensure that the IRB community becomes and stays engaged in conversations about data sharing as they move forward, and becomes and stays integrated into processes that are created to manage and share sensitive data? More ambitiously, how can we build lasting partnerships among the IRB, funding, and publishing communities around these issues?

**2:30-2:45      Coffee break**

**2:45-3:30      Tensions and Solutions from a Researcher’s Perspective**  
**QDR Facilitator: Colin Elman**

In this penultimate session, we consider the processes and rules we have been discussing from the researcher perspective. Researchers are responding to a variety of cues and have a variety of incentives (intellectual, professional, etc.). Our goals in the short term are to help scholars (1) who conducted their research in advance of the social sciences’ reconsideration of data sharing and research transparency to navigate new calls for openness; and (2) who are conducting research now to consider how best to balance the imperatives of protecting research participants and sharing data. When and through what means can we interact with and help them? More broadly – how do new rules / changes impact disciplinary value systems, and how will such changes impact researchers’ behavior? What scholarly behaviors do we want to make sure we do not inadvertently incentivize as we think about the interactions of data sharing and protecting human participants?

**3:30-4:15      Debrief and Plan for Next Steps**  
**QDR Facilitator: Diana Kapiszewski**

We will briefly review the workshop discussions, and evaluate the value and feasibility of IRBs, funders, editors, and data repositories continuing the dialogue, and working together to develop frameworks, guidance, and model language for scholars doing research on human subjects, applying for grants for such work, or submitting manuscripts based on such work. If there is a consensus that sustained cross-institutional interaction would be useful, what modalities and platforms might be used – for instance, newsletters, or a listserv, or an extended workshop series, or webinars? What already exists on which we could build? What would be our specific goals?

## Helpful Resources

### General:

Crosas, Mercè, Julian Gautier, Sebastian Karcher, Dessi Kirilova, Gerard Otalora, and Abigail Schwartz. 2018. “Data Policies of Highly-ranked Social Science Journals”. *SocArXiv*. March 30. <https://doi.org/10.17605/OSF.IO/9H7AY>

Neylon Cameron. 2017. “Compliance Culture or Culture Change? The role of funders in improving data management and sharing practice amongst researchers.” *Research Ideas and Outcomes* 3: e21705. <https://doi.org/10.3897/rio.3.e21705>

### Sample IRB guidance:

Cornell IRB guidance for researchers from the social/behavioral sciences (downloadable file) <https://www.irb.cornell.edu/documents/IRB%20consent%20template%20-%20social-behavioral.doc>

### Guidance from repositories on responsible data sharing and differential access:

Dataverse (Harvard University)

<https://dataverse.org/best-practices/harvard-dataverse-general-terms-use> (section on User Uploads > Restrictions; esp. 6.)

ICPSR (University of Michigan)

<https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter1.html> (section on Informed Consent under Ethics and Privacy)

<https://www.icpsr.umich.edu/icpsrweb/content/ICPSR/access/restricted/index.html>

- Specific federally-sponsored collection example:  
<https://www.icpsr.umich.edu/icpsrweb/content/NACJD/restricted.html>
- Specific state-based census collection example:  
<https://community.isr.umich.edu/public/mcrdc/MCRDCOverview/DisclosurePrevention.aspx>

Odum Institute (University of North Carolina, Chapel Hill)

[http://odum.unc.edu/files/2017/05/Policy\\_UNCDataverseTermsofUse\\_20170501.pdf](http://odum.unc.edu/files/2017/05/Policy_UNCDataverseTermsofUse_20170501.pdf)

[http://odum.unc.edu/files/2017/05/Guidelines\\_DataSecurity\\_20170501.pdf](http://odum.unc.edu/files/2017/05/Guidelines_DataSecurity_20170501.pdf)

Qualitative Data Repository (Syracuse University)

<https://qdr.syr.edu/guidance/human-participants>

<https://qdr.syr.edu/guidance/human-participants/informed-consent>

<https://qdr.syr.edu/termsandconditions> (Section 6. Responsible Use of Data and Data Confidentiality)

<https://qdr.syr.edu/guidance/human-participants/access-controls>