Working with Sensitive Research Data:

The Qualitative Data Repository's WSRD Project

Diana Kapiszewski (QDR and Georgetown University)
Dessi Kirilova (QDR)
Sebastian Karcher (QDR)
IASSIST
May 31, 2019

QDR

QUALITATIVE DATA
REPOSITORY

The Place of Data in Scholarship

- *Data underlie all empirical knowledge claims made by scholars*

- *Not new, but newly recognized, that sharing data*
    - *Allows them to be used for secondary analysis*
    - *Facilitates research transparency and strengthens research both methodologically and substantively*
    - *Can supplement teaching*

Recent Drivers of Increased Data Sharing

- *Data sharing = increasingly an imperative for researchers*
- *Funder and publisher requirements – enforcers*
- *Technological and infrastructure advances – enablers*
- *Researchers – implementers*
  - *Sometimes reluctant*
  - *Increasingly recognizing how data sharing benefit them*
- *Sometimes research participants themselves!*

However…

- With sensitive human participants data, sharing imperative can conflict with ethical and legal responsibilities to protect human participants

- Tension particularly relevant for social science

- Some believe more acute for qualitative data, given their closer relationship to the social world from which they are drawn

- Hence, IRBs* as an additional agent of great importance in the data sharing landscape

*We use IRB here as a shorthand for ethics boards more broadly

The underappreciated DMP – IRB nexus

- *IRBs' and data repositories' goals overlap*
  - *IRBs keep human participants safe*
  - *Data repositories keep data derived from interactions with human participants safe (and thus also keep those participants safe)*
- *Key point of interaction:* **informed consent scripts**
  - *Guide interaction with human participants and whether/which data can be shared*
- *E.g. in US, 45CFR46.111 criteria for IRB approval of research already list a lot of DM-related topics, but the exact plans by PIs are not always well-thought out/spelled out in their IRB protocols*

The Data Repository-IRB Nexus

- *Observation from QDR's experiences with consent scripts:*
  - *researchers do not always recall what they promised to human participants.*

- *Original Q from QDR's experiences with consent scripts:*
  - *What do IRBs suggest to researchers with regard to soliciting informed consent?*

Working With Sensitive Research Data (WSRD)

- *Broader project focused on facilitating the responsible sharing of sensitive research data*

- *What can be accomplished?*
  - *Bring together parallel conversations*
  - *Increase stakeholders' awareness of and sensitivity to each other's perceptions and concerns.*
  - *Facilitate development of systematic procedures for identifying/dealing with sensitive data*
  - *Promote the introduction of more nuanced, harmonized guidance in stakeholders' worldviews and workflows.*

Current NSF Grant

- Proposal in response to 2018 DCL on "Advancing Long-term Reuse of Scientific Data"

- Collaboration between ICPSR and QDR

- Two-year project (2018-2020) to:
    - Foster enhanced coordination between data repositories and IRBs
    - Align better guidance offered for dealing with sensitive human participant data with practices and capabilities of repos
    - Reveal IRB-DMP connection when not clear

Building on Progress to Date

- 2016-2018 series of three workshops with IRB personnel and others (social scientists, journals, funders, repos)
  - *Incremental broaching of the conversation*
  - *Acceptance of the idea that data sharing can be done ethically and in compliance with IRB regulations*

- 2017 empirical study of IRB guidance from 50 IRBs at leading universities with the most SBE NSF funding
  - *Finding: Little consideration of data sharing*

- Consensus texts for more nuanced options on informed consent as output of May 2018 workshop
  - *Being endorsed by IRB staff*
  - *Hope to have it inform revised guidance materials used by IRBs to educate researchers*

Current work – Phases 1 & 2

- *Conducted 15 individual interviews with senior IRB personnel at a variety of institutions across the country*

- *Holding 2 focus groups with same type participants*
  - *DC on 4/26 and Chicago 5/31*

- *Facilitating 3 workshops for NSF-funded PIs at AERA (Education), APSA (Political Science), computational social science conference*
  - *To hear out what researchers are facing when trying to satisfy the dual mandates (data sharing and human participant protection)*

- *Concluding workshop at PRIM&R SBE "Advancing Ethical Research" conference – to bring findings back to broader IRB community*

Next stage – Phase 3

- *Draw on feedback we are receiving on draft consensus texts*
- *Generate further model guidance, improve consent scripts*
- *Create glossary of terms relating to generation and sharing of sensitive human participant data*
- *Disseminate materials through network channels we've already developed with critical stakeholders*
- *Start designing a potential DataPro tool for IRB protocol recommendations on responsible sharing and reuse*

Early Findings

- *Dramatic change in IRB attitudes and some practices the past 2 years since documentary study*

- *IRBs at R1 universities in particular are clearly aware of the data-sharing imperative and do not intrinsically oppose such activities*

- *Often work with campus data security policies in mind (sometimes in close collaboration with other campus units such as library or data center)*

- *IRBs do not typically create their own clearer guidance on the topic for researchers however*

- *Rarely familiar with data repositories and virtually never interact with them on the front end of data collection (sometimes DUAs sent to IRBs for secondary work)*

Auspicious Moment for Change?

- *Ongoing focus on formal data sharing*

- *Revisions to Common Rule came into force this year and are provoking conversations and encouraging change on many campuses*
  - *Ex: Cornell and Harvard's IRBs*

- *Capitalize on moment to engage on broad coordination on processes for sharing human participants data*

Different communities, intersecting objectives

- *Data management community*
- *IRB community*

"In SBE research, it seems like the most important thing an IRB can do is ensure that the informed consent process and content affords for good decision making by prospective participants."

(From an August 2017 PRIM&R IRB Forum discussion)

Bridging the Ethics-Data Sharing Divide5

- *The ethical and legal component remains primary for all research involving human participants (via PI, IRB process)*

- *Funders, but also individual institutions, might require formal DMPs*

- *Effective research data management is vital to managing risk (e.g., data loss/corruption, inability to validate research, potential for privacy breaches) and the potential for ethical and legal data publication and sharing*

- *Reuse reduces societal cost and participant fatigue (being over-studied)*

A First Step: Informed Consent

- *Elements of a consent script*
  - *New for US: Starts with short, plain-language summary*
  - *Outlines details of the proposed interaction*
  - *Discusses risk / benefits of participation*
  - *Includes mechanisms for withdrawal*
  - *Discusses all intended purposes for the data*
  - *Discusses how data will be managed and the steps that will be taken to keep data safe*
- *Ideally make research easily understood and avoid excessive warnings*

Informed Consent for Sharing Data

Items to look for in informed consent script (to ensure **informed** consent)

- What data human participants are willing to have shared
- What the plans for sharing are—when, where, how, with whom
- For what purposes might shared data be used
- What steps will be taken to keep the data safe
- Easy to understand (appropriate language) and no excessive warnings

Repository-specific roles and solutions

- *Technology-assisted (e.g. access control)*
  - *Timed embargos*
  - *Virtual/physical enclaves*
  - *Secure/encrypted downloads*

- *Policy-based*
  - *User agreements for both depositor and end-user of the data*

- *Workflow-based*
  - *Assistance by repository staff in data collection methods decision-making and DM planning*
    - *Including de-identification strategies, although actual implementation done by PIs*
  - *Disclosure review before publication*

Model Consent Script Language

- *Developed out of previous workshop with IRBs and domain repositories; endorsed by a number of individual IRB professionals*

- *Commitment to the principle of more nuanced arrangements for access*

- *Illustrative texts: guidance and template language for researchers*

- *Typically best to make data sharing opt-in*

- *Three possible scenarios*
  - *(A) data will be de-identified*
  - *(B) full de-ID may not be possible/desirable*
  - *(C) de-identification is not necessary*

Consent Script Language (A)

For use when data will be de-identified

De-identified data generated from the information you provide in our interaction may be shared with the research community (most likely in digital form via the internet) to advance scholarly knowledge. I plan to deposit the data at REPOSITORY X, or at a similar social science domain repository. I will use my best efforts to remove or code (e.g., reference as "Participant #1") personal information that could identify you before the data are shared in an effort to ensure that, by current scientific standards and known methods, no one will be able to identify you from the shared data. Despite these measures, I cannot guarantee complete anonymity.

In summary…

- *Using:*
  - *informed consent and*
  - *data sharing mediated by repositories*

- *Researchers can*
  - *reduce the risks associated with collecting and sharing sensitive data*
  - *share data that might not otherwise not have been shared.*

- *IRBs and repositories can help them do it right*

- *These "more nuanced options" for keeping sensitive data safe are helping the scholarly world move toward optimizing the balance between sharing research data and protecting human participants.*

Question

- *WSRD is mainly US focused. How is the link between RDM and ethics boards elsewhere?*
- *Are there best practices we could learn from?*