

Shifting the Consensus on Sharing Sensitive Research Data

**IASSIST
31 May 2018**

**Colin Elman (Syracuse / QDR)
Diana Kapiszewski (Georgetown / QDR)
Dessi Kirilova (QDR)**

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

Drivers of 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

The Data Repository-IRB Nexus (I)

- 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)
- Important to consider how repositories and IRBs can work together
- One point of interaction: informed consent scripts
 - Guide interaction with human participants AND which data can be shared

The Data Repository-IRB Nexus (II)

- Observation from QDR's experiences with consent scripts: researchers do not always recall what they promised to human participants.
- Question from QDR's experiences with consent scripts: what do IRBs suggest to researchers with regard to soliciting informed consent?

Empirical Study of IRB Guidance

- CASE SELECTION
 - 50 institutions that received the most funding from the National Science Foundation's Directorate for Social, Behavioral & Economic Sciences (FY 2016)
 - *Source:* <https://dellweb.bfa.nsf.gov/Top50Inst2/default.asp> (NSF's awards database)
- RATIONALE:
 - Scholars at these institutions clearly face both calls for data sharing (as per 2011 NSF data-management requirement) and IRB constraints (as per federal regulation that impacts all research in the U.S.)


Empirical Study: Cases

- Almost exclusively large, R1 institutions
 - Set trends and create models
 - Are in position to lead re-consideration of IRB practices
- Three-quarters of the institutions in our sample have a medical school
- 66% of institutions in our sample are public

Empirical Study: Materials

- Different models, different materials
 - For schools with combined IRBs: combined documentation
 - For schools with separate IRBs for SBE (most)” documents associated with SBE IRB
- For 4 of 50 schools, no publicly available documents on websites → added next 4 schools
- 114 items
- Schools vary in terms of # of guidance documents
 - Range 1-5; most commonly 2-3
 - Most of our figures are percentages of all *documents*, not institutions

Analysis: Our Project in Dedoose



dedoose
Great Research Made Easy

IRB guidance language on data sharing | [Logout](#) | [Account](#)

[Home](#) [Codes](#) [Media](#) [Excerpts](#) [Descriptors](#) [Analyze](#) [Memos](#) [Training](#) [Security](#) [Data Set](#) [Back](#) [Projects](#)

Project: IRB guidance language on data sharing

Users: 1
Media: 114
Descriptors: 114
Excerpts: 26
Codes: 16
Code Applications: 29

[Import Data](#) Spreadsheets, Documents, Audio, Video, Projects, etc.
[Export Data](#) Excerpts, Media, Codes, Descriptors, Project, etc.

Media

Type	Title	Added	User
	Boston U_Consent Template_August_2015.docx	06/07/2017	Dessito
	Berkeley_Consent Form_Sample_MediaRecordsl	06/07/2017	Dessito
	Berkeley_Consent Form-Template_SocBehav.do	06/07/2017	Dessito
	Berkeley_Consent Form-Sample_Interview_Audic	06/07/2017	Dessito
	Boston U_Video_Consent_Form_March_2012.do	06/07/2017	Dessito
	Carnegie Mellon University-IRB-consent-7.20.201	06/07/2017	Dessito
	Case Western Reserve_Informed Consent Docun	06/07/2017	Dessito
	Columbia_ConsentTemplate.doc	06/07/2017	Dessito
	Columbia_Recording of Human Subjects Policy	06/07/2017	Dessito

Codes x Descriptor

Field: University Type

administrative guidance on IC

Private 100.0%
Public 0.0%

benefits

Private 100.0%
Public 0.0%

benefits to society or scientific knowledge

Private 100.0%

Codes

- administrative guidance on IC
- benefits
 - benefits to society or scientific knowle...
- data sharing
 - can't guarantee anonymity
 - de-identified data sharing only
- focus groups
- mandatory guidance
- privacy / confidentiality / data security

Excerpts: 26

Resource Johns Hopkins_ConsentTemplate.doc Added 09/08/2017 # Codes 2

RISKS/DISCOMFORTS:
Describe any reasonably foreseeable risks and discomforts to the participant.
If appropriate, include the following statement:...

Resource Johns Hopkins_ConsentTemplate.doc Added 09/08/2017 # Codes 1

Most sections are required for all consent forms.

Resource Boston U_Consent Template_August Added 09/08/2017 # Codes 2

Are there any benefits from being in this research study?

Descriptor Ratios

University Type

Public
Private

Empirical Study: Initial Analysis

Descriptive: simple searches and counts of words/phrases

- *Data*
- *Data sharing, data access, data availability*
- *Repository (data repo, institutional repo), archive, storage*
- *Transparency*
- *Access controls, access conditions*
- *Anonymize, de-identify*

Empirical Study: Initial Findings

- “Data” mentioned in 66% of documents
- “Access” mentioned in 54% of documents
- “Anonymous” in 18%
- “Data sharing” in 3 ITEMS = 0.02%
- “Repository” in 4 ITEMS + “Archive” in 5 ITEMS = 0.07% (combined for both terms)
- “Transparency”, “Access controls” / “Access conditions”, “Reuse” / “re-use”, “Secondary use” / “Secondary data” → ZERO hits!
- Continued analysis of the data

Empirical Study: Tentative Conclusions

- Some IRBs have yet to consider how IRB regulations and data sharing interact
- IRB offer little guidance to researchers on data sharing
- Lots of room for collaboration so we can
 - Better serve researchers, who are now under a dual mandate
 - Better protect human participants.

Continuing Work

- Broader project focused on facilitating the responsible sharing of sensitive research data
 - Initial goal: coordination between data repos and IRBs
 - Broader goal: discussion among multiple stakeholders
- 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
 - Facilitate the introduction of more nuanced, harmonized guidance in stakeholders' worldviews and workflows.
- Likely neither quick nor simple!

Auspicious Moment for Change?

- Renewed focus on data sharing
- Revisions to Common Rule should come into force soon and are provoking conversations and encouraging change on many campuses
 - Exs: Cornell and Harvard's IRBs
- Capitalize on moment to engage on broad coordination on processes for sharing human participants data

Types of Change

		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

Encouraging Change

- Basis: unappreciated synergies between Data Management Plans (DMPs) and IRB protocols
- DMPs help scholars consider how the data will be managed and potentially shared
- Decisions made when developing DMPs interact with those made when developing IRB application
- Clear overlaps between data repositories and IRBs
- ERGO: utility of repository / IRB collaboration!

A First Step: Informed Consent

- Informed consent = a researcher's key tool/responsibility
- Elements of a consent script
 - 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 Data Sharing

- How can informed consent be secured in ways that enable data sharing?
- Requests for archiving / sharing / other future uses should be explicit
- Items to discuss with project participants
 - What data can be shared?
 - How will data be shared - when, where, how, with whom?
 - For what purposes might shared data be used?
 - What steps will be taken to keep the data safe?
- ***Informed*** consent
- Research repercussions?

Model Guidance and Consent Script Language

- Developed out of previous workshop with IRBs and domain repositories.
- Commitment to principle of more nuanced arrangements for access.
- Illustrative texts: guidance and template language for researchers → endorsement of principle of more nuanced options for data sharing.
- 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.

Consent script language (B)

For use when full de-identification might not be possible or desirable

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. Due to the nature of the information, full de-identification of those data might not be possible. I plan to deposit the data at REPOSITORY X, or at a similar social science domain repository. Your data will be made available under XXX access conditions. Despite my taking these measures it is not possible to predict how those who access the data will use them.

Consent Script Language (C)

For use when de-identification is not necessary

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. We have discussed the benefits and risks of sharing the data and you agree that the data may be shared without de-identification or other protective measures.

De-Identification Strategies To Decrease Disclosure Risks

- Don't collect identifying information
- Remove direct *and* indirect identifiers Not collecting or removing direct identifiers
- Reduce precision / detail through aggregation
- Generalize meaning of detailed variables
- Restrict upper or lower ranges to conceal outliers
- Combine values into broader categories in narrative
- Best done by researcher him/herself
- Concerns

Repository-Based Solutions for Sharing Data

- Workflow-based solutions
 - Repository staff assist potential depositors
 - Repository staff conduct disclosure reviews
- Technology-based solutions
 - Timed embargos
 - Secure/encrypted downloads
 - Enclaves (virtual / physical)
- Policy-based solutions
 - Depositor and user agreements

In Summary...

- With
 - Two key tools...
 - Help of the broader data services and ethics communities...
 - Employing strategies just discussed...
- Scholars can devise solutions and strategies for
 - Reducing the risks that can be associated with collecting and sharing sensitive data
 - Making shareable data that might not otherwise not have been shared.
- These = “more nuanced options” for keeping sensitive data safe - helping us move toward optimizing the balance between sharing res data and protecting HPs.

THANK YOU!