



Lessons from working with sensitive data...

Presenter:

Amy Pienta, Ph.D.

Director of Business & Collection Dev.

ICPSR Research Scientist

Presentation for:

Research Data Alliance P17

Date: April 22, 2021

IG: Sensitive Data

National Addiction & HIV Data Archive Program

- Funded under a contract with the National Institute on Drug Abuse (NIDA) since Fall 2009
- Infrastructure to facilitate sharing of sensitive data on addiction and/or HIV research data
- Technical assistance to investigators and their staff to prepare data for archiving
- Technical assistance and training to data users

A Case for Sharing Sensitive Data

Sensitive data can be shared and there are many benefits that outweigh the risks.

Research participants, especially, want their data to be used for research when their privacy is maintained (Damschroder et al. 2007; Fiesler and Proferes 2018).

Sensitive data can often be de-identified and shared in a straightforward manner.

Data in NAHDAP

- Public-use data sets
 - Contain no identifiers so minimal disclosure risk
 - Downloadable, no IRB approval/exemption
 - Must agree to ICPSR standard Terms of Use
- Restricted-use data sets
 - Contain indirect identifiers, have disclosure risk
 - Documentation downloadable, need IRB approval/exemption
 - Requires Restricted Data Use Agreement
 - Obtain data by secure download or VDE



Example of Sensitive Data Being Shared

Population Assessment of Tobacco and Health

A collaboration between the National Institutes of Health (NIH) and the Food and Drug Administration (FDA), the study was launched in 2011 and started the first wave of data collection in 2013.

The PATH Study was one of the first tobacco projects that NIH and FDA have worked on together since Congress gave FDA authority to regulate tobacco products. The study looks at tobacco use and how it affects the health of people in the U.S.

What makes PATH sensitive?

Political and industry interest in the outcomes of tobacco regulatory research

Biomarkers

Breadth and depth of information on tobacco use (esp children and teens)

Parents and children in the study

How ICPSR Disseminates PATH

ICPSR operates a FISMA Moderate secure enclave to hold and access restricted PATH data

Early PATH data releases are restricted before fully de-identified public-use data are released.

Data are made available to users in a VDE so that staff vet all output to ensure minimum thresholds are maintained.

Data are made available to academic researchers, but also to commercial organizations able to meet our standard for engaging in ethical research.

And, a bit of the unexpected...

An instance of an intentional data breach led us to create a training for VDE users to ensure that they understand processes for working with the data in the VDE.

The newly required training surfaced multiple instances of teams that had failed to follow the data security plan for working in the VDE.

The training and string of breaches also sensitized archive staff to noticing and reporting minor protocol violations and there are frequent data security incidents.