

# Restricted Data Contracts: Current & Future Directions

C3: Ethical Sharing & Management of Data

**Lisa Broniszewski**

Data Compliance Specialist  
Population Research Institute  
Penn State

**IRB**

**Jen Darragh**

Senior Research Data  
Management Consultant  
Duke University Libraries

**Data Protection  
Plans**

**Lisa Neidert**

Assistant Research Scientist  
Population Studies Center  
University of Michigan

**Data  
Management**

## AGU's DMM<sup>SM</sup> Best Practices and the Data Lifecycle



### Practices Spanning Entire Lifecycle:

- Measurement & Analysis
- Process Management
- Process Quality Assurance
- Risk Management
- Configuration Management
- Sustainability, Consistency & Resilience





# IRB: Current & Future Trends

LISA BRONISZEWSKI  
DATA COMPLIANCE SPECIALIST  
POPULATION RESEARCH INSTITUTE | THE PENNSYLVANIA STATE UNIVERSITY  
IASSIST 2017, SESSION C3, ETHICAL SHARING AND MANAGEMENT OF DATA

# IRB - Current

- ▶ Why get? Journals need/require a determination to publish
- ▶ DUAs that require:
  - ▶ Most DUAs have “de-identified” data
    - ▶ Fall under expedited review
    - ▶ HIPAA 18 list – ask researchers to check off variables they will have access to
    - ▶ PSU – special protocol for HSR-Chart Review and Analysis of Existing Restricted Data Set Study – they made questions specific to existing data
- ▶ HSR Training - CITI

# IRB - Future

- ▶ NIH Changes
  - ▶ “Clinical Trail” definition expands
  - ▶ Must register on [clinicaltrials.gov](https://clinicaltrials.gov) – takes hours to complete
  - ▶ New training required
  - ▶ Not retroactive
  - ▶ IRB of record (file in primary institution)
- ▶ Common Rule Revisions – January 2018
  - ▶ Exclusions (+)
  - ▶ Exemptions (-)





# The INS and OUTS of Data Protection Plans

JEN DARRAGH  
SENIOR RESEARCH DATA MANAGEMENT CONSULTANT  
DUKE UNIVERSITY LIBRARIES

IASSIST 2017, SESSION C3, ETHICAL SHARING AND MANAGEMENT OF DATA

# What is a Data Protection Plans (DPP)?

- ▶ An additional document submitted with the restricted-use data request describing measures to protect both **physical** and **electronic** security.
  - ▶ You are required to fill this out for local hosting of the data at your institution.
- ▶ Some use a fillable form with suggested and/or recommended practices/environments, others are far less prescriptive and open-ended.
- ▶ Some may require your system to meet a national-level standard - e.g. NIST SP 800 -171 starting January 2018 (link: [NIST SP 800-171](#))
- ▶ Can be the ultimate determining factor as to whether you can obtain the data or not.

# Elements of a DPP – Physical Security

- ▶ Where is original physical media stored? What building and room (with address), cabinet, safe, etc.
- ▶ How is physical access to the data secured? Mitigated entry to building, locked room, cage for CPU, etc.?
- ▶ What is the physical location of the researchers when they are working with the data?
- ▶ How is access to printed output controlled? Location of printer, is it networked, is it shared, how are printouts stored and disposed of?



# Elements of a DPP – Electronic Security

- ▶ Stand-alone desktop computers, closed LAN, dedicated servers
- ▶ Hardware and software, including firewall and virus protection
- ▶ Full disk encryption
- ▶ Address removable or portable media (including laptops)
- ▶ What is included and excluded during system backups
- ▶ Who has system-level access for tech-support
- ▶ Security breach protocols
- ▶ And.....

# Other things to consider

- ▶ If the physical environment of the equipment or the researcher changes, you will need to get approval from the provider before the equipment can move OR the researcher regains access to the data.
- ▶ Institutional-level policies around the storage and access of restricted-use secondary data may not exist.
- ▶ You might have to write policies.
- ▶ You need to track any changes to physical and electronic environments and clear them with the provider as they instruct (some care more than others).
- ▶ The DPP is a part of a whole. Proper restricted-data lifecycle management is a must.



# Data Use Agreements: The good, the bad and the ugly

LISA NEIDERT

ASSISTANT RESEARCH SCIENTIST

INSTITUTE FOR SOCIAL RESEARCH | UNIVERSITY OF MICHIGAN

IASIST 2017, SESSION C3, ETHICAL SHARING AND MANAGEMENT OF DATA



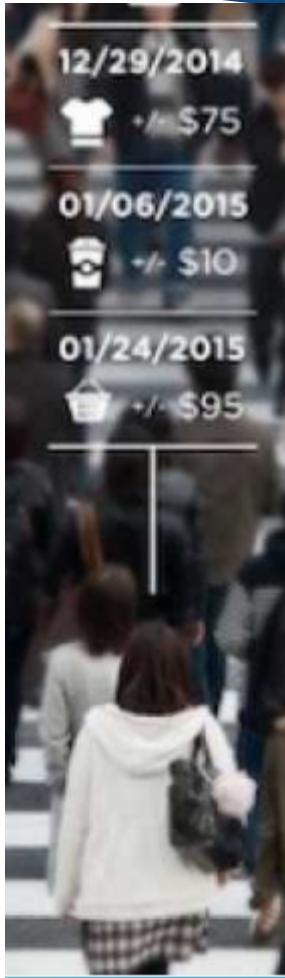
# Opening Questions

- ▶ Do they conflict with OPEN data?
- ▶ If DUAS are necessary, why are some so bad?
- ▶ Solution to the heterogeneity

# OPEN Data

- ▶ DUAs serve a purpose
  - ▶ Necessary for sensitive, potentially identifiable data
- ▶ Re-identification of respondents
  - ▶ Lower response rate
  - ▶ Reluctant respondents
  - ▶ Non-response on sensitive questions
- ▶ Re-identification examples in the national media

# False Alarm or Re-identification?



Your Anonymous Credit Card Data Isn't So Back To You

Your Anonymous Anonymous

## Not so Anonymous: Four Bits of Credit Card Metadata Might Give Away Your Identity

According to the study carried out by Massachusetts Institute of Technology (MIT) researchers, four bits of credit card metadata is capable of re-identifying nearly 90 percent of the people.

A recent report in *Science Magazine* revealed the soft underbelly of what was once considered a well-armored use of “anonymized” consumer information. The study’s authors were able to successfully identify consumers based on several anonymized data sets—specifically, their credit card purchases.





# A Critique of DUAs

- ▶ DUAs vary across data providers
  - ▶ Conditions
  - ▶ Order in which conditions are presented
  - ▶ Some details are not easy to find
- ▶ Formatting
  - ▶ Section Headings are nice
  - ▶ No footnotes!
- ▶ Need a way to keep track of this complexity

100



	Academic														Bio						Federal												
	Add Health	ANES	CHNS	DNORS	Frag-Fam	HRS	IPUMSI	LA FANS	Mt. Laurel	NDA	NIS	NSFH	PSID	RLMS-HSE	ADNI	dbGaP	FHS	MESA	NHLBI	TCGA	WHI	BLS-NLS	CMS	NCES	NDI	NSF	NVSS	AIR: PT	ANRF	High Scope	IC		
Data Use																																	
Cannot merge new files w/o amendment		x	x		x	x		x	x		x	x										x	x			x	x	x					
Data Limitations or Errors															x		x			x									x				
Data used for statistical purposes	x	x		x	x	x		x	x	x	x		x									x		x	x	x	x	x	x		x		
Hazards: Bio-markers, genomes																	x			x													
No administrative uses of data		x			x	x			x		x													x		x			x				
No re-identification	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x					x	x				
No redistribution of data	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x		x	x	x	x	x	x	x	x	x	x	x	
Only minimum necessary data																							x										
Stick to original research plan		x	x			x		x	x		x		x	x		x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	
Training & Awareness				x												x				x		x				x							
Work files to be destroyed periodically	x																																
Publication																																	
Cite/acknowledge data	x	x	x	x	x		x	x		x	x		x	x	x	x	x	x	x	x	x	x	x				x			x			
Disclaimer language														x	x	x	x					x				x							
Disclosure conditions	x	x	x	x		x		x	x		x	x	x						x	x		x	x	x			x						
Intellectual Property																				x													
Joint ownership of publication										x																						x	
Make effort to publish																x	x	x	x			x	x										
Mutual Benefit:Share variables/give talk																																x	
No case listings in publications	x		x				x		x															x		x							
PMCID	x												x					x															
Pre-review required						x									x		x	x				x	x		x	x						x	
Public Posting																	x			x													
Send articles		x																															
Inspections & Penalties																																	

# Uses of the Conditions database

- ▶ Create Structured DUA for each data provider
  - ▶ Conditions are in the same order for each DUA
  - ▶ Sorted in order relevant to researchers:
    - ▶ Data Use | Publication | Penalty | Security | Legal | Administrative
- ▶ Uses
  - ▶ Education
  - ▶ Information
  - ▶ Analysis



# Education: First Use & Annual Reminder

All of our restricted data are governed by conditions of use or a data use agreement (DUA). You need to be aware of the conditions and follow them. The full language of the DUA can be found here:

<http://www.cpc.unc.edu/projects/addhealth/contracts/RestrictedUseContractNew.pdf>

An alternative is to look at the DUA in a structured format, where the conditions are sorted by category with the conditions that concern user behavior in a more prominent location. All PSC DUAs are structured the same way:

<http://www.psc.isr.umich.edu/dis/data/dc/prv-cncls.html?i=1147>

And, finally, here are the conditions you should pay particular attention to:

**Data Contract Language for:  
The National Longitudinal Study of Adolescent to Adult Health**

May 8, 2017, 1:05 pm

## **Data Use**

### **Data used for statistical purposes**

V. A. That the Data Files will be used solely for statistical analyses: that no attempt will be made to identify specific individuals, families, households, schools, institutions, or geographic locations not provided by Add Health; and that no list of Sensitive Data at the individual or

# Education: **Highlight** unusual conditions

## **Work files to be destroyed periodically**

Referenced in the data protection plan. [Workfiles need to be destroyed and re-generated every 6 months.

3. Only one complete copy of the Add Health data is permitted; however, time-delimited temporary data analysis files may be created. Temporary data analysis file(s) must be deleted every six months and recreated, as necessary, to complete analysis. Temporary data analysis files should be deleted upon completion of a project.

## **PMCID**

V. K. That all journal articles based on analysis of Confidential Data from Add Health receive a PubMed Central reference number (PMCID). Journal articles must be submitted to PubMed Central to receive a PMCID. The method of PubMed Central submission and Investigator responsibility for submission depend on the journal and journal publisher:

# Minimal Conditions

- ▶ No Re-distribution
- ▶ Statistical uses only
- ▶ No Re-identification
- ▶ Inadvertent Disclosure
- ▶ Disclosure Conditions
  - ▶ Data specific
- ▶ Expiration Date
- ▶ Media [Return | Destroy]



# Final Thoughts

- ▶ DUAs are necessary
- ▶ Data Providers should examine their DUAs
  - ▶ Compare against others via conditions database [information]
  - ▶ Data protection plans get more attention than DUAs
- ▶ Need norms for user behavior
  - ▶ Hard to establish with broadening of field of social science
    - ▶ Computer science, data science
  - ▶ Legislate? [Gelman]



## Data Use Agreement [DUA] Portal\*

Use our database to compare contracts to see if they share the same conditions; to create easy to read reports of contract conditions for your contract users; or to pick and choose conditions for a new restricted data contract.

### Quick Comparison

This grid shows all contracts x all conditions. An x in the grid indicates that this contract has language relevant to that condition. Click on the x to see the contract language. Click on the [Data Provider](#) name at the column headings to get the data contract conditions in a simplified structured report for that data provider. Click on the [Conditions](#) in the row headings to see which contracts require this condition (including the language).

Note, go to the [contracts](#) link for the actual contracts.

### Contracts

These are links to replicas of the contracts or letters of agreement held by researchers at the Population Studies Center. Note that the numbering system for each contract relates to the notation and language in the Quick Comparison view.

### Inspections

These are links to inspection checklists. These are pretty typical of an inspection experience - maybe not all in a single visit, but over time these questions will be asked. Take a look so you can be prepared.

### References

These are links to presentations based on the contracts database.

\*This website content is drawn from the [PSC Data Services Data Contracts database](#).

## SEARCH DATA SERVICES

### Meta Search

search across all the following databases:

### Data Catalog

Data and documentation

### KnowledgeBase

Common questions and answers.

### Resources

Entire collection of data resources.

## LATEST DATA NEWS

Census Bureau director resigns

In the nick of time?

The title says it all

Anything you can't measure you can't manage

New word of the day: frugging

# Invitations

- ▶ Please help us to benchmark what is happening at other institutions by completing a short survey at <https://www.surveymonkey.com/r/XXF9S98>
- ▶ Please join us at the **Birds of a Feather** (BoF) session to further discuss these topics – Thursday afternoon