

## *Disclosure Review at the National Addiction and HIV Data Archive Program at ICPSR*



Amy Pienta and  
Kaye Marz, ICPSR





**Why Did ICPSR Build this Capacity?**



## Restricted Data at ICPSR

- Bill and Melinda Gates Foundation funded research project - Measures of Effective Teaching Longitudinal Database
- Classroom videos and survey data
- Virtual Data Enclave (VDE) was developed for these data
- Used by many ICPSR projects today.

## VDE and NAHDAP

- With the VDE established, we proposed to the National Institute on Drug Abuse to use the VDE to provide access to sensitive, addiction data.
- Data added by NAHDAP tend to be longitudinal, often are regional, and also contain sensitive information about drug use, risky behaviors, HIV status.

## Tobacco Regulatory Research

- Food and Drug Administration began in 2009.
- Family Smoking Prevention and Tobacco Control Act
  - Regulate tobacco products
  - Need for data and new research to guide regulation



## The PATH Study

- Population Assessment of Tobacco and Health
- National, longitudinal
- Tobacco->Health (+ smokeless, e-cigarettes)
- Survey of 46,000 people ages 12+
- Goal provide broad public access, including to the tobacco industry, protect confidentiality
- Restricted & Public Use Data (biospecimens)





# The VDE Process at NAHDAP

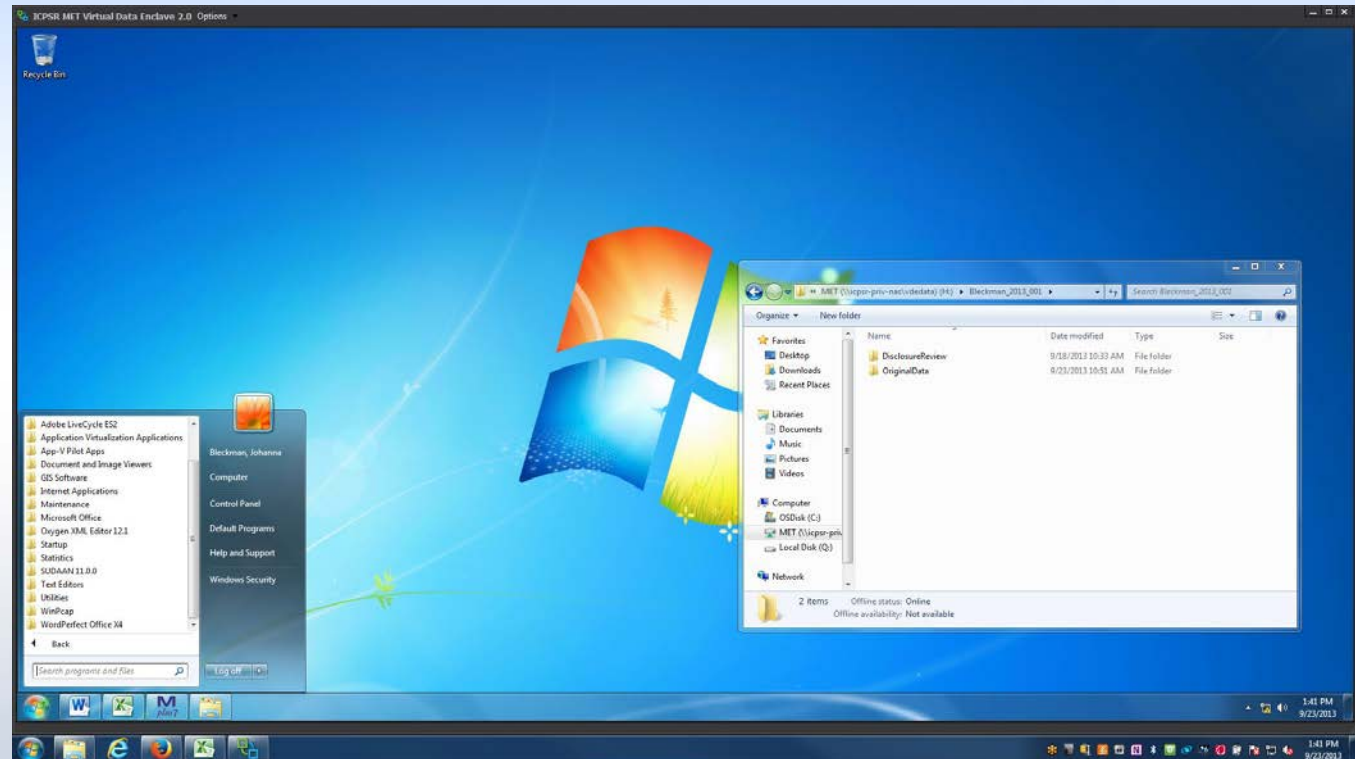
## Virtual Data Enclave

- Virtual machine launched from your desktop machine
- Functions just like your local machine, but with restrictions on what can enter and exit the environment
- Full suite of statistical packages and other software



## Virtual Data enclave (VDE) – user experience

VDE



My laptop's task bar

## NAHDAP Volume of VDE Applications

- We have 106 initiated applications, 55 approved.
- We review and give feedback within 72 hours.
- Delays occur to process modified DUAs.



## Common Reasons for Delays for Access

- Revise description to identify datasets needed
- ID an investigator with experience with confidential data with 1-3 references
- Require separate DUAs for people at different institutions
- CV/biosketches do not provide up-to-date information on affiliation with the institution.

## How we keep application to VDE “Easy”

- Informative Interface (Create Project) walks applicants through steps to make a request.
- Text appears below boxes in the application with further guidance.
- Automatic emails provide instructions for that step.
- Staff available (via email) to answer Qs



## VDE Guide

Our VDE Guide is a series of web pages and a downloadable PDF



**NAHDAP** Kevin Marx Logout

National Addiction & HIV Data Archive Program

ABOUT US ▼ FIND DATA FIND VARIABLES DEPOSIT DATA PUBLICATIONS TRAINING CONTACT US HELP

ABOUT THE VDE MAKING A REQUEST ▼ USING THE VDE ▼ OBTAINING VDE OUTPUT ▼

**Guide to Data Access Using the ICPSR Virtual Data Enclave (VDE)**

[View this guide as a PDF](#)

**Why are NAHDAP data in the ICPSR VDE?**

The National Addiction & HIV Data Archive Program (NAHDAP), funded by the National Institute on Drug Abuse (NIDA), is hosted at the Inter-university Consortium for Political and Social Research (ICPSR). NAHDAP data in the ICPSR Virtual Data Enclave (VDE) are restricted from general dissemination to protect the confidentiality of the individuals and/or organizations represented in the data. To access data in the VDE, a Restricted Data Use Agreement (RDUA) must be established between the University of Michigan and the researcher's institution.

**How does it work?**

A link to the online application is provided in the Access Notes of the online study description for the restricted-use dataset. Researchers interested in accessing the restricted-use data using the VDE must complete the online application. After the researcher creates an online project, the RDUA is sent by email to the researcher. The researcher returns the signed RDUA to NAHDAP by email. **NOTE:** Researchers on the same project but at different institutions must each provide a signed RDUA—one from each institution—because this agreement is a legal document with the researcher's institution, not the researcher here at ICPSR.

The Guide will help you understand about the VDE, how to apply for access, how to use the VDE, and about obtaining output from the VDE.

## Volume of NAHDAP Disclosure Reviews

- Project manager has processed 47 disclosure reviews across 15 projects since January 2016.
- About 1/3 are returned for the output to be modified in order to clear or more information on what is to be reviewed.



## Common Problems

- Not clearing the sample size and minimum cell size threshold.
- More information on what files are to review/be received outside the VDE.
- The reviewer being able to understand the characteristics of the sample and procedures in order to conduct the review.

## Cell Count Too Small

|           | Years of experience | 0 to 1 | 2   | 3   | 4   | 5   | 10  | 15  | 20  | 25  | 30 | 35 | 40 | Total |
|-----------|---------------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|-------|
|           | Sample size         | 110    | 166 | 183 | 143 | 133 | 450 | 280 | 174 | 117 | 97 | 58 | 19 | 1930  |
| Ethnicity | Black               | 5      | 8   | 8   | 11  | 9   | 24  | 14  | 12  | 9   | 4  | 3  | 1  | 108   |
|           | Hispanic            | 35     | 55  | 65  | 62  | 51  | 128 | 78  | 50  | 24  | 21 | 12 | 2  | 583   |
|           | White, other        | 60     | 91  | 95  | 80  | 81  | 298 | 188 | 112 | 84  | 72 | 58 | 20 | 1239  |
| Gener     | Male                | 22     | 35  | 37  | 25  | 20  | 90  | 56  | 38  | 24  | 18 | 10 | 4  | 379   |
|           | Female              | 88     | 131 | 141 | 123 | 113 | 360 | 224 | 130 | 99  | 78 | 49 | 15 | 1551  |

|           | Years of experience | 0 to 1 | 2   | 3   | 4   | 5   | 10  | 15  | 20  | 25  | 30 | 35+ | Total |
|-----------|---------------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|----|-----|-------|
|           | Sample size         | 110    | 166 | 183 | 143 | 133 | 450 | 280 | 174 | 117 | 97 | 77  | 2007  |
| Ethnicity | Black or Hispanic   | 40     | 63  | 73  | 73  | 60  | 152 | 92  | 62  | 33  | 25 | 18  | 691   |
|           | White, other        | 60     | 91  | 95  | 80  | 81  | 298 | 188 | 112 | 84  | 72 | 78  | 1317  |
| Gener     | Male                | 22     | 35  | 37  | 25  | 20  | 90  | 56  | 38  | 24  | 18 | 14  | 393   |
|           | Female              | 88     | 131 | 141 | 123 | 113 | 360 | 224 | 130 | 99  | 78 | 64  | 1615  |



## Wait Times

- Users are told to expect 10 days for the review.
- Currently the average number of days for submission to receipt of cleared files is 8 days.
- Requests returned have the longest turnaround time are where the researcher asked to resubmit.
- Times from receipt to 1st return is ~3 days

## Attempts at Efficiencies?

- De-centralized, NAHDAP staff (2 people) do NAHDAP reviews.
- We only review the files in the disclosure review folder.
- Require users to point to the syntax that produced table.
- Prefer to review tables/slides, not output files.



## Clearance Memo

- A clearance memo describing the review process guides users in preparing their files for review.
- Added this to “Output Review Folder”