

## SYNTHETIC DATA PROJECT:

*EXPANDING MLDS DATA ACCESS AND  
RESEARCH CAPACITY WITH SYNTHETIC DATA*

Michael E. Woolley, Ph.D., Associate Professor  
Laura M. Stapleton, Ph.D., Associate Professor  
University of Maryland

## OUTLINE

- Applications of Synthetic Data
- From proposal to funded project
- Central project tasks and benefits to Center
- How do *Synthetic Data* work?

2

## APPLICATIONS OF SYNTHETIC DATA

- Several government programs—U.S. and others—have synthetic data strategies
- The U.S. Census has a publically available synthetic data set for the Security of Income and Program Participation (SIPP) data
- No state longitudinal education data system has investigated using a synthetic data strategy

3

## SLDS 2015

- Synthetic Data Project designed and proposal written by Research Branch
- Proposal forwarded to MSDE for inclusion in SLDS 2015 submission
- Awarded Funding by IES (2.6 million)
- Includes: expert consultation, project coordinator, and database manager
- MOU being negotiated, funding expected to be made available by the State in April

4

### CENTRAL TASK 1 – 2016-17

- Build Database
- Will include three Center data sets
  - K12 to PS
  - K12 to WF
  - PS to WF

*Benefit to Center*

- These three data sets will make some of our research analyses and dashboard work more efficient

5

### CENTRAL TASK 2 – 2017-19

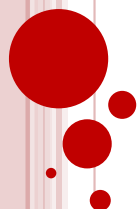
- Build synthetic versions of the three data sets
- Test security and safety of those synthetic data sets
- Test the research utility of those data sets

*Benefits to Center*

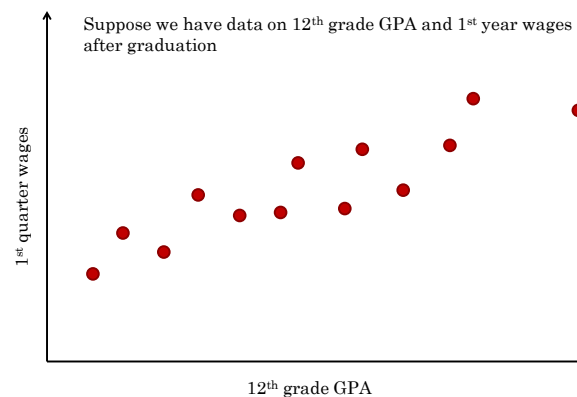
- The Center will have developed synthetic data procedures, then empirical evidence of the security and utility of those synthetic data sets
- Center will have synthetic data for response to data requests

6

### HOW DO SYNTHETIC DATA WORK?



### CENTER DATA WITH TWO VARIABLES



8

