

ACI 1443014 (PI Reiter):

Main objectives

- Develop techniques and open source software that government agencies and data stewards could use to provide access to confidential social science data
- Create ways for researchers, students, and others to access record-level data on work histories of federal employees, provided by the U. S. Office of Personnel Management (OPM)

Overarching strategy

- Develop an integrated system including
 - unrestricted access to fully synthetic data, coupled with
 - means for approved researchers to access the confidential data via remote access solutions, glued together by
 - verification servers that allow users to assess the quality of their inferences with the synthetic data so as to be more efficient with their use (if necessary) of the remote access to the confidential data.

What have we done so far?

- Fully synthetic OPM data, including synthetic careers
- Verification measures for Level 1 (low trust) users that satisfy **differential privacy**
 - Plots of residuals versus predicted values for regression
 - ROC curves in logistic regression
 - Signed significance of regression coefficients
- Remote access granted to approved researchers from multiple universities without Duke credentials

Challenges and Directions

- Sustainability
 - Satisfying OPM (and Duke) so they want to sustain it
 - What if OPM pulls the plug?
- Scaling up the infrastructure
 - Expand to users from outside trusted group
- Technical
 - “Privacy budgets” in the context of verification