

Looking for Answers to Questions We Don't Know Yet



Peter S. Meyer

National Center for Health Statistics
Division of Research and Methodology

Board of Scientific Counselors
December 5, 2018

Remote Access Analysis Tools for Working with Government Data

Problems

- **Data Science**
 - Methods
 - People
- **Big Data**
 - Storage
 - Processing

Some Questions We Know

- **How are we going to develop a Data Science research program?**
 - Discover appropriate methods
 - Non-probabilty samples
 - Grow our own people
- **What tools are we going to use to make Big Data work?**
 - CDC, HHS, Census, NCI, Dell, Niter and Brock, Hortonworks, Amazon, NIOSH
 - Own the data not the platform

One Promising Answer

- **Oak Ridge National Laboratory**
 - Health Data Institute
 - CADES
 - Data Scientists
 - SUMMIT
 - Qualifications
 - NCI – genetic
 - VA – EHRs
 - Georgetown – health data linkage
 - DOE

Future Technology

- **Differential Privacy**
- **Synthetic data**
- **Query system**
- **Remote execution**
- **Machine Learning**