NCHS Virtual Data Enclave (VDE) Project Overview

J. Neil Russell, Ph.D.

Director, Research Data Center

BSC Presentation

May 26, 2022

For Discussion and Feedback

- The VDE has an increased risk of disclosure compared to a Research Data Center (RDC):
  - Do the benefits of having a VDE for NCHS and researchers outweigh the additional risk?
- NCHS was approved for ten virtual machines. Demand may exceed supply:
  - Is a VDE the future of accessing restricted-use data?
  - Should NCHS proceed with expanding VDE services?
- The operational cost of the VDE is unknown. Researchers will incur fees:
  - Should the VDE fee structure be substantially different from the RDC fee structure?

GOAL: Expand Data Access

The Evidence Act* directs federal statistical agencies to expand access to data assets – current NCHS data access modes are:

- **Open data assets**: free access to public-use microdata files

- **Web-based data query system**: free access to tool to create tabular data output via web-based interface

- **Physical data enclave**: access to restricted-use microdata files via NCHS Research Data Center (RDC) and Federal Statistical Research Data Centers (FSRDC)

- ***NEW* Virtual data enclave (VDE):** remote access to restricted-use microdata files

* 44 U.S.C. 3582(a)

Virtual Data Enclave (VDE)

- Allows an approved researcher to log into a secure, remote server to access and analyze restricted-use data from a remote location.

- The researcher does not travel to a physical data enclave (i.e., RDC) for data access.

- The researcher accesses restricted-use data from an approved "off-site" location - usually their employer's worksite.

- Designed to prevent unauthorized disclosure of restricted-use data.

VDE Benefits

- Adds a tier of data access as per the Evidence Act.*

- Operates almost 24/7 – data access is not limited by **pandemics,** extreme weather, or civil unrest.

- Allows researcher flexibility in managing their project schedule – not limited to business days/times.

- Decreased cost to researchers, e.g., reduced travel, no background investigation:

  - Could result in expanding the pool of researchers

- Uses existing National Institute for Occupational Safety and Health (NIOSH) VDE enterprise services.

* See 44 U.S.C. 3563 and 44 U.S.C. 3582(a)

VDE Project Schedule

- Nov 2020:   Approved for CDC Data Modernization Initiative (DMI) funds

- June 2021:   Project approval

- Sept 2021:   Project kick-off

- April 2022:   NCHS Director approval for operation

- July 2022:   Onboard pilot CDC projects

- Fall 2022:   Onboard other projects

VDE IT System Security

- The NCHS VDE will be hosted on NIOSH servers in Atlanta CDC HQ.

- NIOSH servers have an authority to operate (ATO) at the moderate level.

- Restricted-use data encrypted at rest on NIOSH servers.

- NCHS data has its own volume on NIOSH servers – no commingled data.

- Uses encrypted via virtual private network (VPN) – connection between NIOSH server and researcher's computer is encrypted.

- Researchers use two factor authentication for login.

- NIOSH employees and contractors will be CIPSEA* agents.

- VDE system logs available to monitor researcher activity.

* Confidential Information Protection and Statistical Efficiency Act (CIPSEA) of 2018 (Title III of the Evidence Act; 44 U.S.C. 3561-3583)

VDE Security

- No download or copy/paste of data or output

- No printing available

- No Internet access

- No email access

- Certain Windows programs deleted (e.g., Snipping Tool, Command prompt, Remote Desktop, Internet browsers and PowerShell)

- Disclosure risk controlled through Data Use Agreement (DUA) between researcher's employer and NCHS:

  – DUA specifies data protection requirements

  – e.g., no photos or recording of information displayed in VDE

- NCHS staff will review researcher output for disclosures before release

VDE Project Challenges

- Additional risk of disclosure compared with RDC:
  - The DUA will prohibit certain researcher behavior, but ability to monitor behavior is reduced when compared to using an RDC
- Demand may be greater than supply – NCHS "bought" ten virtual machines from NIOSH for concurrent operation
  - May need to schedule researcher use of machines
- True operational cost of VDE is unknown – researchers will incur fees
- Researchers must "learn" a new IT system