

National Center for Health Statistics

Peter S. Meyer

“Virtual Data Access” for
Statistical and Research
Purposes

2014 FCSM Statistical Policy Seminar
December 15-16, Washington D.C.



Overview of NCHS/RDC/ANDRE

- **ANDRE was first implemented on 4/8/1997**
 - The system was reengineered in 2007 and is currently being updated
- **NCHS goal for creating a virtual data access system**
 - Allow access to researchers who could not travel to Hyattsville
 - Decrease the cost of using NCHS data
 - Legislative mandate for making data available confidentially
- **NCHS 'buy-in' for this virtual data access system was achieved:**
 - Presented to Senior Staff and they agreed
 - Truly novel idea

Development of ANDRE – *Brief Description*

- **ANDRE works by:**
 - Remote execution
 - Authenticate > Scan > Execute > Scan > Transmit Approved Output
- **ANDRE is maintained by:**
 - RDC and CDC IT staff
- **ANDRE employs:**
 - Procedure management
 - Cell suppression

Development of ANDRE – *Resources/Challenges*

- **ANDRE requires certain resources:**
 - A yearly budget of ~\$30,000
 - 2.5 FTEs (2 Computer Scientists and .5 Statisticians)
 - General IT support from CDC

- **Challenges faced when developing ANDRE:**
 - A. Getting past the bureaucracy (EPLC, C&A, IR Governance, ...)
 - B. Finding personnel who can do the work

- **These challenges were overcome by:**
 - A. Ignoring them
 - B. Unsolved and pending

Development of ANDRE – *Testing*

- **Testing ANDRE before going live involved:**
 - Communications
 - Computational
 - Penetration

- **Glitches/Potential issues found when testing ANDRE:**
 - A. Queueing of job submissions
 - B. Overflow of data to SAS log space

- **These were corrected by:**
 - A. Time limiting specific runs
 - B. Text editing, i.e., blanking out all logs

Development of ANDRE – *IT & RISKS*

- **IT Challenges faced when developing ANDRE**
 - A. Making all of the software work together
 - B. Version control with SAS and expansion into other software packages
 - C. Creating safeguards against intruders

- **These challenges were mitigated through:**
 - A. Vijay Gambhir
 - B. Move the management of procedures onto a SQL server
 - C. Used the CDC infrastructure for firewalls and malware protection

Conducting Business for ANDRE – COSTS

▪ Charges and costs associated with ANDRE

- There is a charge to use ANDRE for all users
 - Credit of \$3,000 is provided for NCHS sponsors
- Initial setup fee = \$750
- Continuing monthly access fee = \$750
- The fees generate a ~10% cost recovery which is less than the cost recovery for research done in the labs
- The fees provide nonmonetary benefits
 - Social deterrence against malefactors
 - Additional vetting of project in cases where researchers are grantees
 - Align researcher behavior by providing incentives

Conducting Business for ANDRE – *SYSTEM USERS*

- **Most of the researchers using ANDRE come from the academic community**

- **Reactions of users:**
 - They do not initially like using ANDRE
 - The veracity of the output was questioned
 - ANDRE would occasionally “bomb”
 - The error is most often caused by the user not reading the documentation
 - Every new version require about a year of intense trouble shooting
 - Usually, researcher are very satisfied with the experience in the end

Conducting Business for ANDRE – USAGE

- **The number of users was flat for a few years but is currently increasing**
- **Questions regarding data quality are countered by using off the self statistical packages**
- **Users are in constant contact with ANDRE administrators though an email help address**
- **We handle data requests on a case-by-case basis**
 - The number of requests is 5-10 per week
 - The request require about 10% of an FTE
 - The number of request sty constant

Conducting Business for ANDRE – *IMPACT*

- **The benefits of maintaining the system outweigh the cost for main**
- **We are able to make different types of data available**
 - Hosting data from other HHS agencies
 - Genetic data
- **The RDC annual customer questionnaire indicates that we are getting the job done**
- **Ongoing concerns with ANDRE are:**
 - Continued development of in-house IT applications
 - Expanding to other processing platforms

Conducting Business for ANDRE – *MOVING FORWARD*

- **The original vision for the system has been fulfilled**
 - What next?
- **The future may look very different**
 - Differential Privacy
 - Homomorphic Encryption
 - Micro-Grouping
- **Remote Access/Remote Execution is necessary**
 - Provides access efficiently at less cost
 - Secondary data analysis will become more important

Peter S. Meyer
pmeyer1@cdc.gov
rdca@cdc.gov

For more information please contact Centers for Disease Control and Prevention

1600 Clifton Road NE, Atlanta, GA 30333

Telephone: 1-800-CDC-INFO (232-4636)/TTY: 1-888-232-6348

Visit: www.cdc.gov | Contact CDC at: 1-800-CDC-INFO or www.cdc.gov/info

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.