



Statistics Canada
www.statcan.gc.ca



Real Time Remote Access (RTRA)

APDU 2015

September 2015

David.Price@statcan.gc.ca

Real Time Remote Access

- **Online remote access facility allowing users to run, more or less in real time, data analyses on microdata located in a central, secure location**
- Statistical analysis software (SAS) programs that run descriptive statistics
- confidentiality applied **automatically** to outputs
- no primary manual intervention
- Streamlined access process
- ability to interrogate Statistics Canada microdata at any time
- access from any computer with Internet access, using a secure username and password

RTRA features

- Data Quality Indicators for Bootstrap weight Surveys
- 81 Surveys and data sets available through the System
- Several Statistics available through RTRA
 - Frequencies
 - Means
 - Medians
 - Percentiles
 - Ratios
 - Proportions

Issues Around Data Access

- Legal requirements – ***The Statistics Act***
- Increased public awareness of privacy
- Concerns about data security
 - Malicious Data hackers/snoopers
 - Technology
- Increased demand for information
- **Demand for faster Access**
- Sophistication of research
- Record linkage
- Several access programs already offered service

Real Time Remote Access Development



Obstacles to overcome

- Network air gap
- Risk – control the risk or manage the risk
- Contracts and consequences

What's at stake

- STC reputation – Enhance or ruin
- Researchers reputation

Principles of the development of RTRA

- Confidentiality
 - Statistics Act
 - Protect the Microdata
- Costs
 - Limited financial capacity
 - **Re-use** as much as possible
 - Other considerations
 - Maximum use of data
 - New mode of access
 - Risk management

Managing Risk of Disclosure

- Managing the risk of disclosure
 - Application process
 - 2 Contracts
 - Terms and conditions
 - Letter of Agreement with organization sponsoring the analytical work.

Controlling the Risk of Disclosure

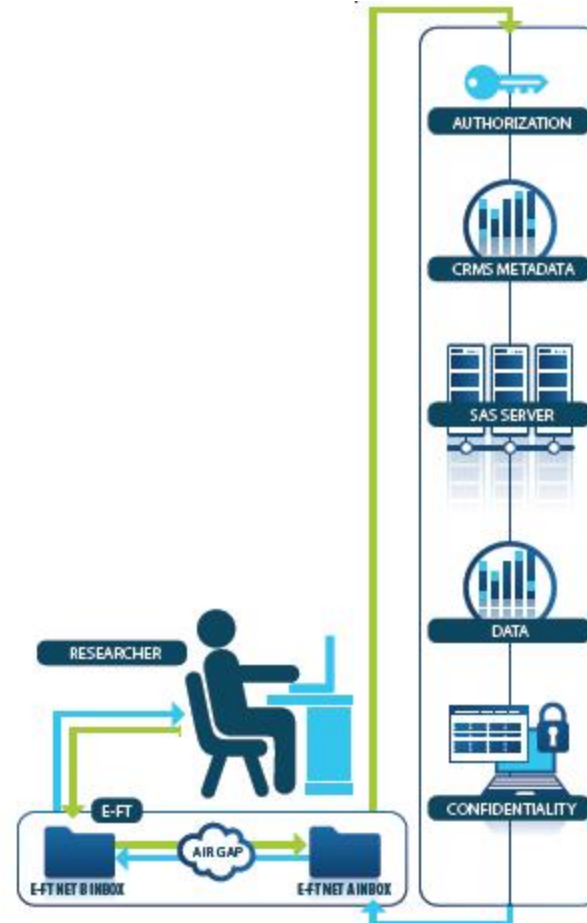
- The microdata files
 - Removal of some variables
 - Small Geographies
 - Potentially sensitive variables.
 - No access to the weights (they are applied automatically)
- Pre-request controls
 - Enabling vs. Disabling functionalities
 - Data Step, proc sort, modified frequency procedure
- Post-request controls
 - Log and Outputs
 - Logs are vetted for confidentiality (no counts returned to user)
 - Custom RTRA log & SAS log returned to user.

Disclosure Control at the program level

- Email notification for account creation temporary password
- Submit SAS program (no SAS support)
 - Maximum of 10 successful submissions, with a maximum of 10 table for each submission per day until midnight
 - Name of the program need to start with TAG name (ex: nlscy2004_anameyouwant.sas)
 - No libname is allowed. Your input dataset will be under RTRA libname
 - Authorized command
 - datastep
 - proc sort
 - Not permitted string: %, &, email, libname, filename, firstobs, FTP, obs, option/options, first., last., ods, _N_, and weight.
 - Use custom procedures to create tables
 - weight will be applied automatically using a proc freq
 - input: datasetin, datasetout and list of variables for the “by” statement

RTRA system process

- RTRA System comprises other corporate systems
 - CRMS
 - EFT
- Pre-Scan & processing in SAS
- Disclosure control post scan via custom program



Disclosure control methods

- Simple and efficient
- Additive and Controlled Rounding (ACROUND)
 - Rounding base is survey-driven
 - Protect against possible link to PUMF
 - Small impact on precision is as close to controlled as possible (i.e. where distances from original values are less than the rounding base's length)
 - All outputs are weighted
- Bound variation and primary suppression for statistics other than frequencies

Output

- Vetting (protection against disclosure)
 - Controlled rounding base
 - Maximum of 5 dimensions
 - Maximum number of cells in a table is 393,216* (includes the totals and sub totals)
 - Maximum of 500 possibilities for a variable, if more no table will be create.

Uses of the RTRA System

- Uses of the RTRA
 - **Projects that don't require complex modeling can be accomplished through the RTRA with descriptive statistics.**
 - RTRA would be suitable when the PUMF data lacks the required detailed but the project is too small or timelines preclude a submission to get access at an RDC.
 - **Access to data sets that don't have a PUMF.**
 - Project feasibility can be determined using the RTRA before finalising and submitting a project proposal to get direct access.

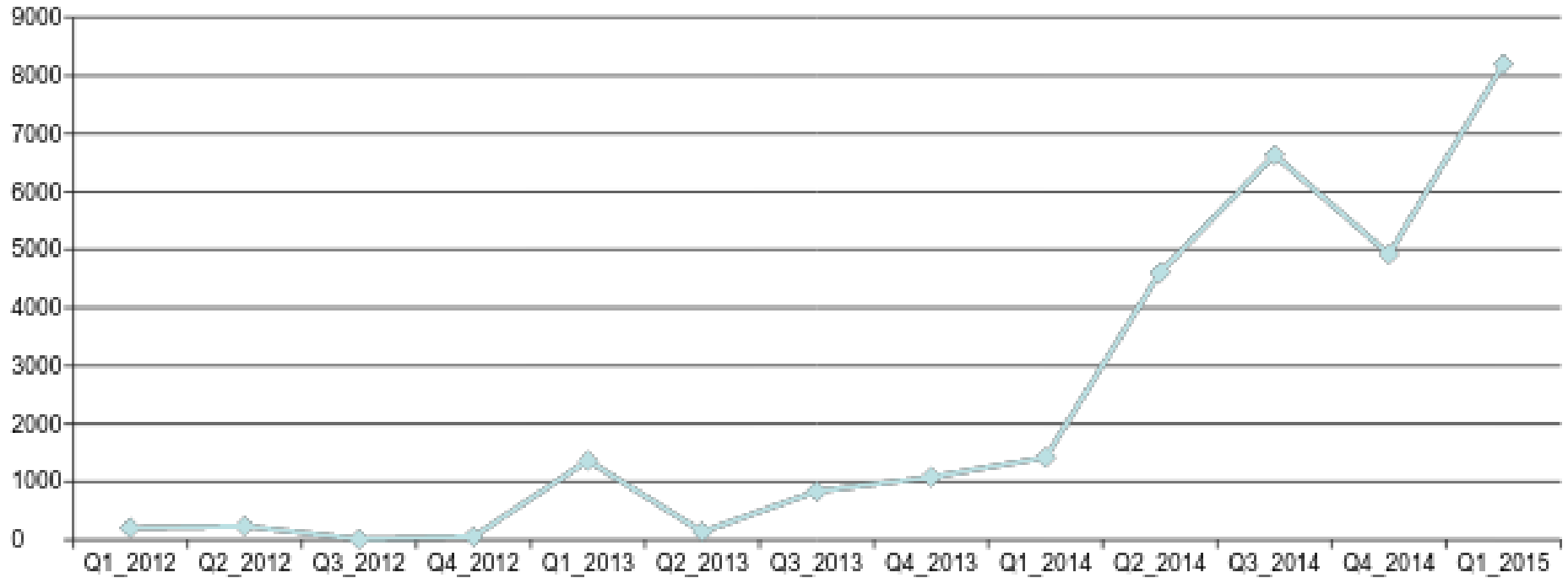
Your Implementation of the RTRA

While the RTRA can provide fast access to a lot of data, a particular set of skills are required.

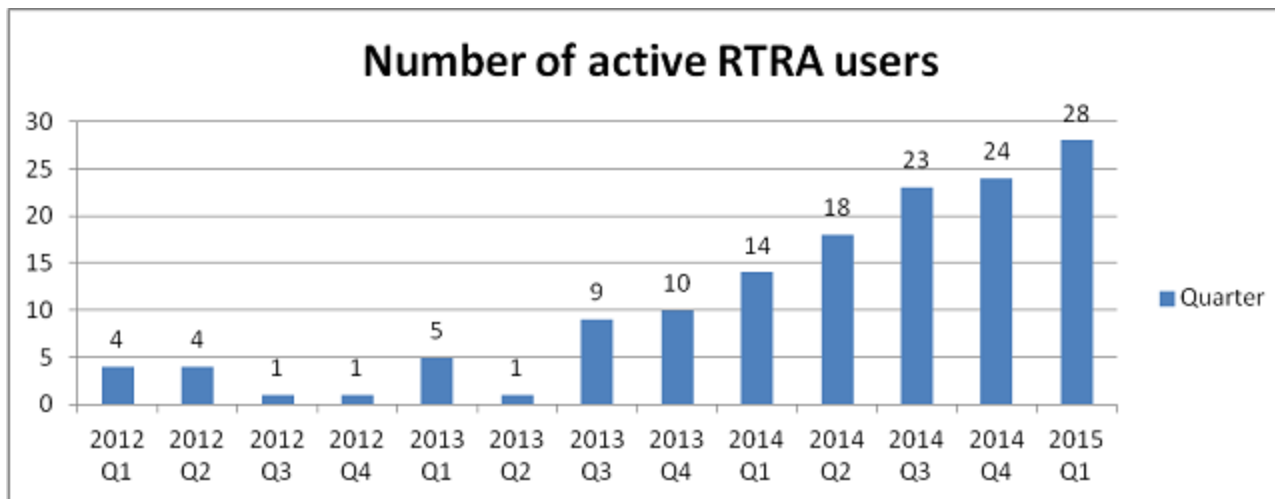
- Capacity to understand and work with microdata
- Basic level SAS programming

We provide extensive training and on-line support for users to get them up and running. Once users are comfortable with the system they become very independent.

RTRA use over time



Users of the RTRA over Time



Users of the RTRA over Time

