Confidentialisation of Data

Australian Government
Statistical Forum
1 November 2012

Melissa Gare
Analytical Services Branch

Overview

- What is confidentiality and why should we care?
- Framework for managing identification risk
- Overview of the ABS suite of data products
- Latest ABS developments in interactive user querying of detailed unit record data
- Current and future research and development directions

Information is power

- Banker in Maryland obtained a list of patients with cancer
  - compared with list of clients with outstanding loans
  - called in the loans of clients with cancer.

Source: Data confidentiality: a review of methods for statistical disclosure limitation and methods for assessing privacy (Statist. Surv. Volume 5 (2011), 1-29.

Confidentiality: What is it & why should you care?

- It's about obligations – legal/ethical
- Aim – protect identity and release useful data

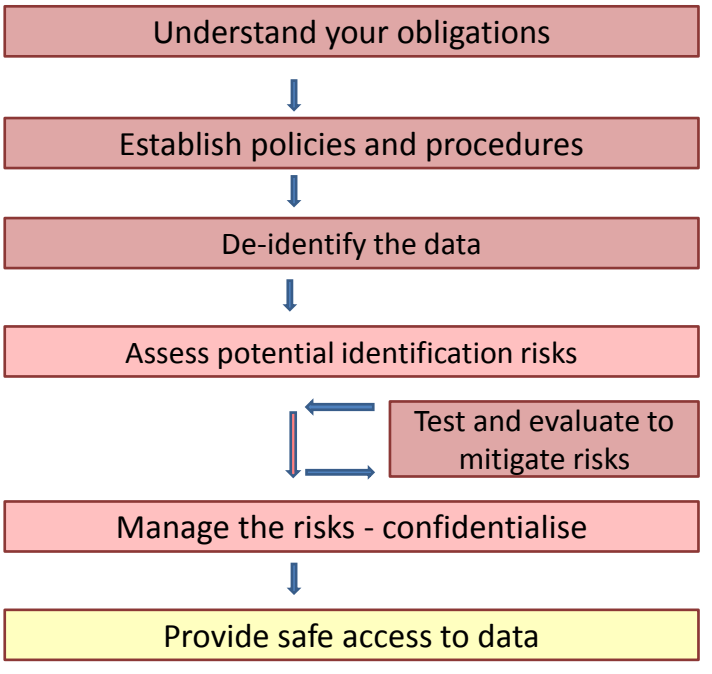
- It's more than removing name & address
- Trust of providers is essential to get good stats

How agencies meet these obligations

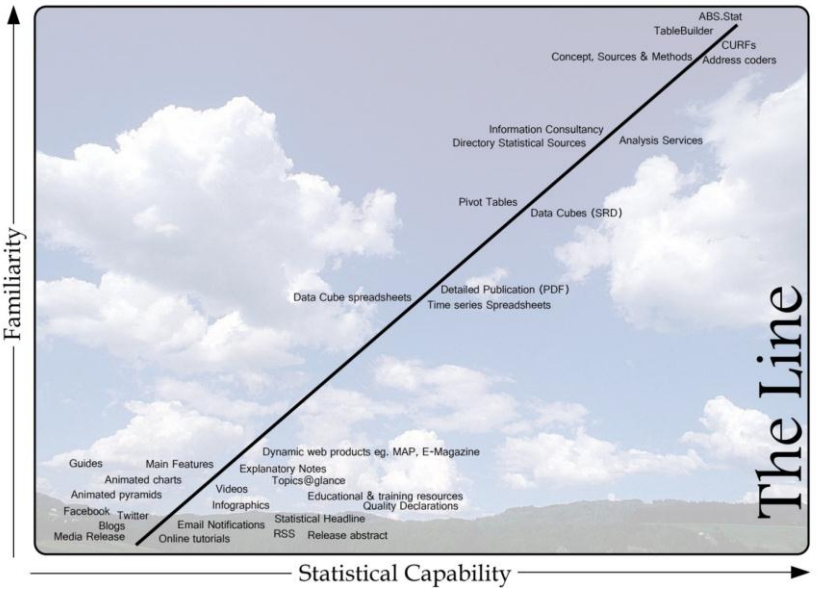
- Implement procedures to address all aspects of data protection

- To ensure that identifiable information:
  - is not released publicly;
  - is available on a 'need to know' basis;
  - can't be derived from disseminated data; and
  - is maintained and accessed securely.

Managing identification risk



ABS Suite of Analysis Products



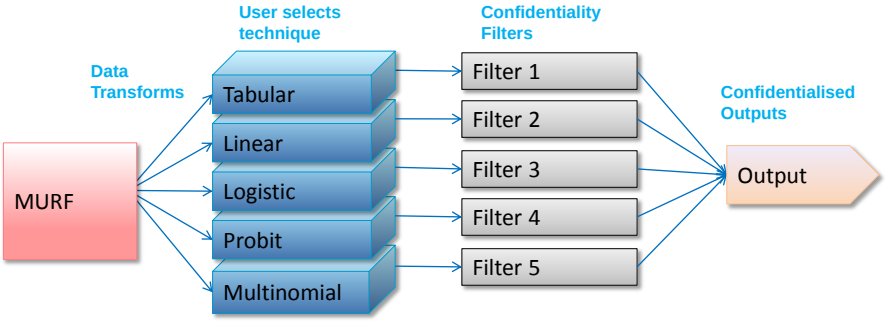
New environment for analysing microdata

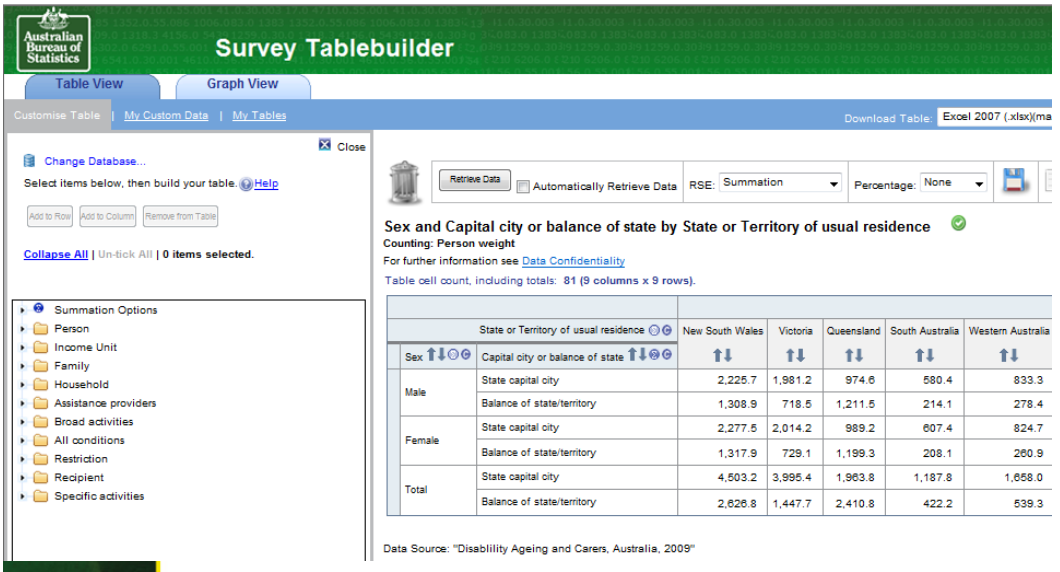
- ✔ Analysis of an expanded and more detailed range of ABS source data with outputs confidentialised rather than inputs
- ✔ User friendly menu driven interfaces
- ✔ Responsive on demand tabulation and analysis
- ✔ Reduced time between publication and when microdata files are available to researchers

- x Reduced set of analytical procedures supported

Current External Researcher Environment

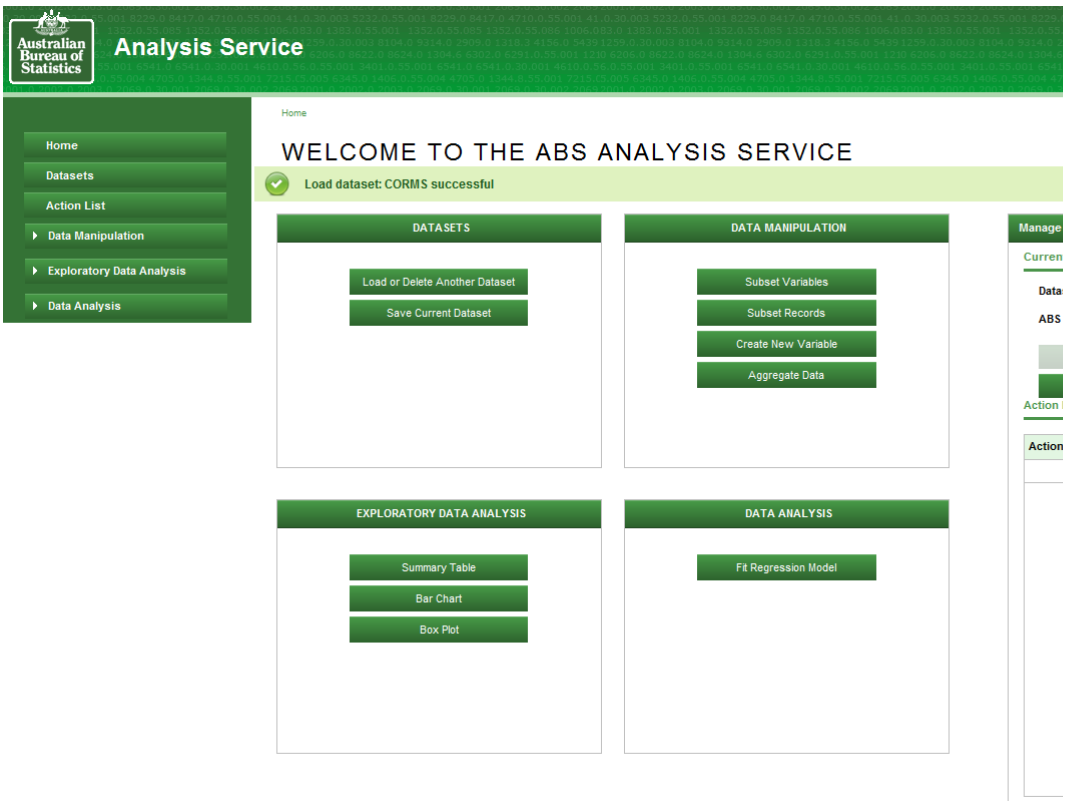
Future ABS Research Environment





•TableBuilder Development

- Census TableBuilder
  - 2006 and 2011 Census of Population and Housing
- Survey TableBuilder
  - Release 1 (population counts)
    - Education and Work, 2011
    - Characteristics of Recent Migrants, 2010
    - Disability, Ageing and Carers, 2009
    - Disability, Ageing and Carers, 2003 (Basic CURF)
  - Late 2012 Release 2 (means, medians, quantiles and custom ranges)
    - 2011-13 Australian Health Survey



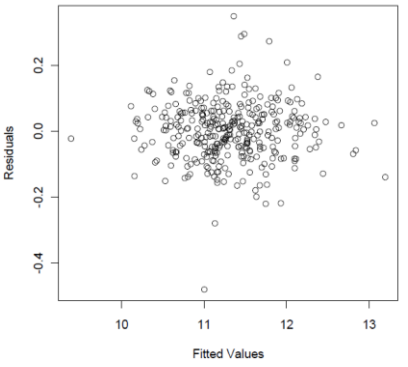


Figure: Standard Residual

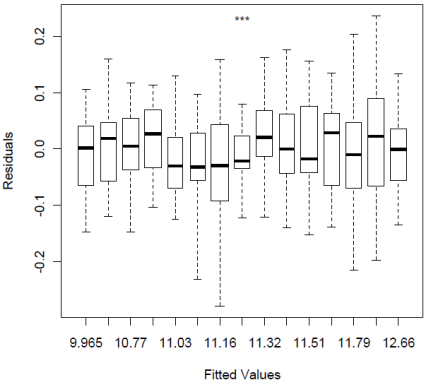


Figure: Confidential Residual Plot

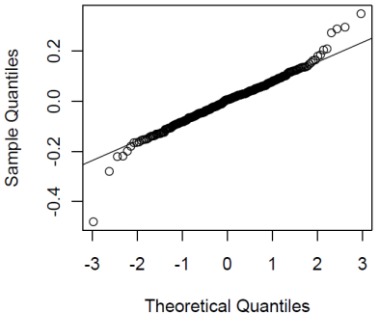


Figure: Standard Q-QPlot

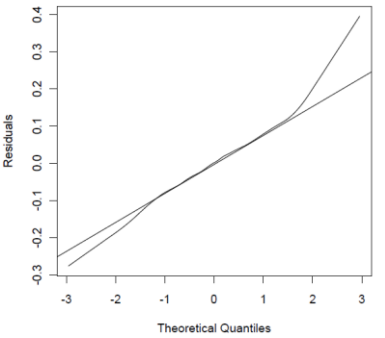


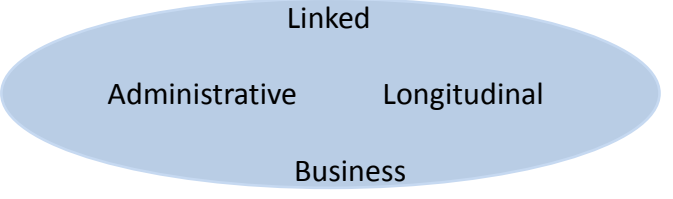
Figure: Confidential Q-Q Plot

•Analysis Service development

- Release 1 – early 2013
  - Dataset manipulation
  - Basic tabulation
  - Modelling (Robust linear, Binomial and Poisson)
- Release 2 – June 2013
  - Enhanced Modelling (Multi-level, multinomial)

Future Research Directions

- Understanding disclosure risks associated with new and/or highly identifiable datasets and emerging analytical techniques
- Developing confidentiality approaches to minimise disclosure risk while maintaining value of these new datasets
- Demonstrating that well developed confidentiality techniques have minimal impact on analyses.



What are AGSF member views on:

- How the ABS can assist the NSS to enhance confidentiality capability and increase availability of data holdings for informed decision making?
- What are the research priorities that would assist policy development?