

# *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



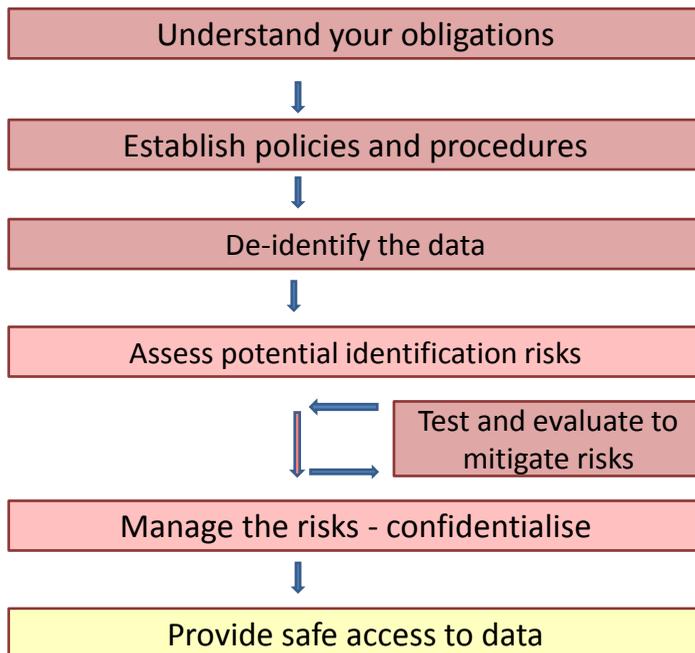


## How agencies meet these obligations

- 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.

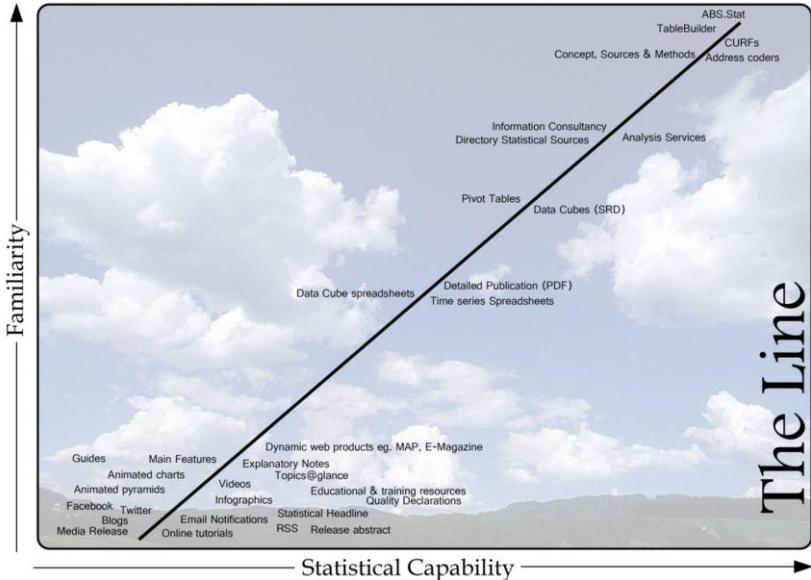


## Managing identification risk





# ABS Suite of Analysis Products



# New environment for analysing microdata

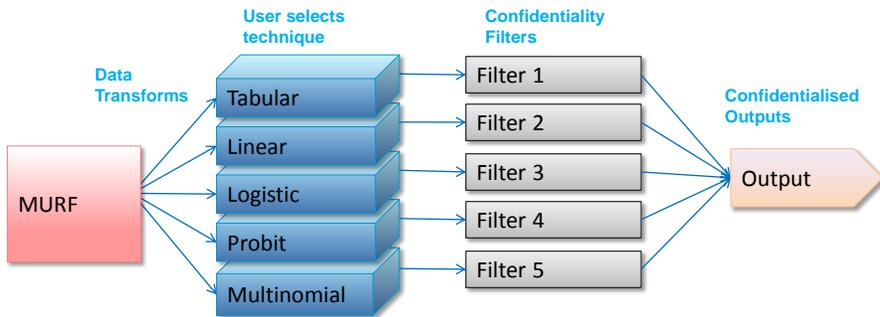
- ✓ Analysis of an expanded and more detailed range of ABS source data with outputs confidentialised rather than inputs
- ✓ User friendly menu driven interfaces
- ✓ Responsive on demand tabulation and analysis
- ✓ Reduced time between publication and when microdata files are available to researchers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



## Current External Researcher Environment



## Future ABS Research Environment



**Survey Tablebuilder**

Table View | Graph View

Customise Table | My Custom Data | My Tables | Download Table: Excel 2007 (.xlsx)|ma

Change Database... | Select items below, then build your table. | Help

Retrieve Data | Automatically Retrieve Data | RSE: Summation | Percentage: None

**Sex and Capital city or balance of state by State or Territory of usual residence**

Counting: Person weight

For further information see [Data Confidentiality](#)

Table cell count, including totals: 81 (9 columns x 9 rows).

		State or Territory of usual residence				
		New South Wales	Victoria	Queensland	South Australia	Western Australia
Sex	Capital city or balance of state					
Male	State capital city	2,225.7	1,981.2	974.6	580.4	833.3
	Balance of state/territory	1,308.9	718.5	1,211.5	214.1	278.4
Female	State capital city	2,277.5	2,014.2	989.2	607.4	824.7
	Balance of state/territory	1,317.9	729.1	1,199.3	208.1	260.9
Total	State capital city	4,503.2	3,995.4	1,963.8	1,187.8	1,658.0
	Balance of state/territory	2,626.8	1,447.7	2,410.8	422.2	539.3

Data Source: "Disability Ageing and Carers, Australia, 2009"



# •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

The screenshot displays the 'Analysis Service' web interface. At the top left is the Australian Bureau of Statistics logo. The main header is a green bar with the text 'Analysis Service'. Below this is a navigation menu on the left with options: Home, Datasets, Action List, Data Manipulation, Exploratory Data Analysis, and Data Analysis. The main content area is titled 'WELCOME TO THE ABS ANALYSIS SERVICE' and shows a success message: 'Load dataset: CORMS successful'. The interface is divided into four main sections: 'DATASETS' (with buttons for 'Load or Delete Another Dataset' and 'Save Current Dataset'), 'DATA MANIPULATION' (with buttons for 'Subset Variables', 'Subset Records', 'Create New Variable', and 'Aggregate Data'), 'EXPLORATORY DATA ANALYSIS' (with buttons for 'Summary Table', 'Bar Chart', and 'Box Plot'), and 'DATA ANALYSIS' (with a button for 'Fit Regression Model'). On the right side, there is a 'Manage' sidebar with options for 'Current', 'Data', 'ABS', 'Action', and another 'Action'.

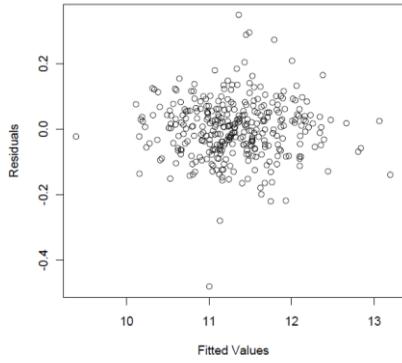


Figure: Standard Residual

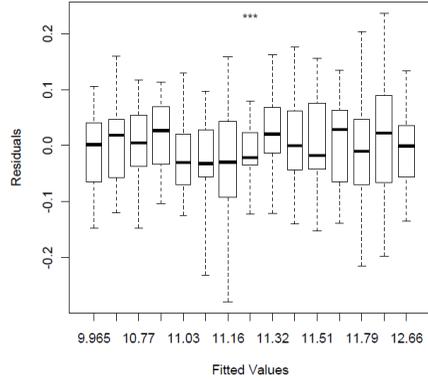


Figure: Confidential Residual Plot

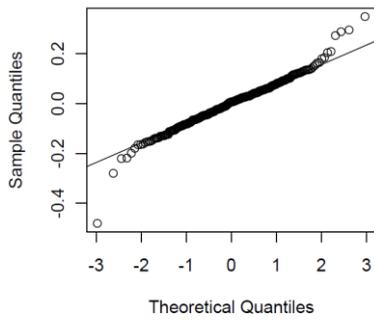


Figure: Standard Q-Q Plot

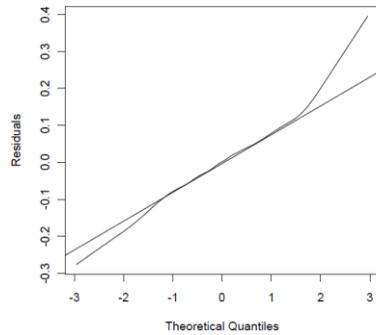


Figure: Confidential Q-Q Plot



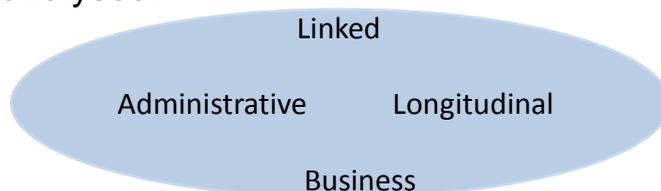
## • Analysis Service development

- 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?

