

Enhancing User Access to microdata in Australia

Australian Government Statistical Forum
November 2010

Melissa Gare

Overview of presentation

- Background
 - ABS Legislation
 - Current modes of analysing Microdata in Australia
- Proposed Future Strategy for remote querying of detailed microdata with automated confidential outputs
 - Survey Table Builder
 - Analysis Service

ABS Mission and Legislation

- We assist and encourage informed decision making, research and discussion within governments and the community, by leading a high quality, objective and responsive national statistical service'
- The Census and Statistics Act through the Statistics Determination **enables** the Statistician to disclose individual statistical records (microdata) for statistical purposes.
 - Identifying information removed
 - Information is **not likely** to enable the identification of person/organisation
 - Given a relevant undertaking

Current Framework for analysis of microdata by researchers

- Researchers' Environment
 - Basic CURFs on CD-ROM
- Remote Execution - RADL
 - Remote access to Basic and Expanded CURFs for statistical analysis in SAS, SPSS and STATA.
- On-site - ABSDL
 - Access to Expanded or Specialist CURFs
- **Special Data Service/Consultancies**

Evaluation of current framework

- ✓ Analysis of Confidentialised URF CD-ROM or RADL
- ✓ Large number of analysis procedures supported remotely in SAS , SPSS or STATA
- ✓ Wide variety of Household survey datasets available for analysis
- x RADL/ABSDDL protections not tight enough to enable analysis of detailed microdata
- x Text file output only (RADL)
- x Have to be able to program in SAS, SPSS or STATA
- x Very few Business and linked CURFs
- x Lengthy CURF creation process
- x Metadata not searchable

Proposed Future Strategy

- Continue to produce Basic CURFs for use in users' own environment
- Look to replace RADL with a new remote environment that enables table generation and analysis to be performed on the detailed data collected
- Enhance ABSDL to enable complex analysis of detailed data including linked and longitudinal datasets.

User defined table generation for Population Census data

- TableBuilder - released in August 2009
- Enables user specified tables (via web interface) to be dynamically produced and confidentialised from the full underlying Census URF → general output
- Very detailed classification levels (including geography)
- Output table dynamically calculated and confidentialised using a perturbation technique that also protects against differencing.

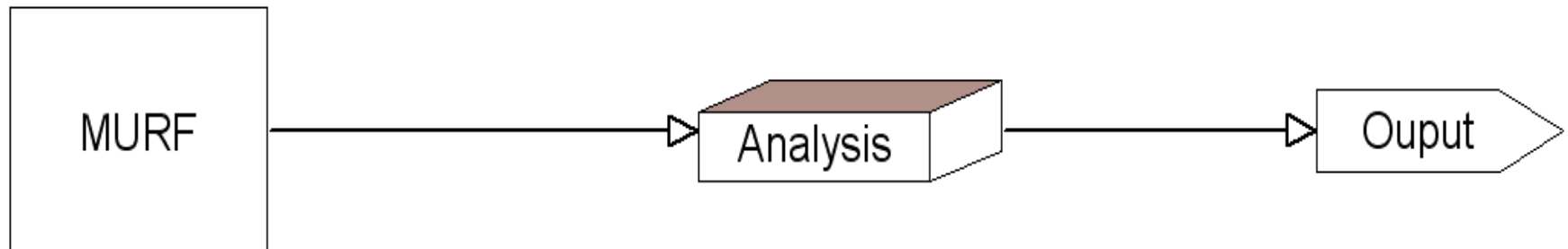
Staged approach to extending TableBuilder to Surveys

- Count estimates from Household surveys (early 2011)
 - International data and metadata standards
 - 'Save-as' SDMX
 - SDMX web services (machine to machine query)
 - Input data specified in DDI format
- Key outputs from continuous household survey data items (late 2011)
- Key outputs from business survey data, linked datasets and administrative datasets, searchable and richer metadata (timing not determined at this stage)

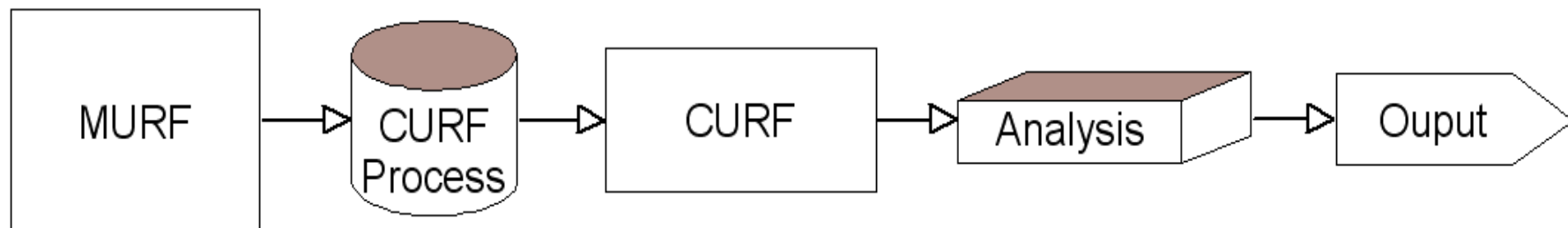
Future Analysis Server

- Service to enable analyses such as modelling (including diagnostics) to be undertaken remotely in real time on detailed data → general output.
- Requires development of automated methods/tailored outputs (filters) to ensure that respondents are not likely to be identified (to comply with ABS legislation)
- Internationally there is increasing interest from researchers and practitioners to design such automated services

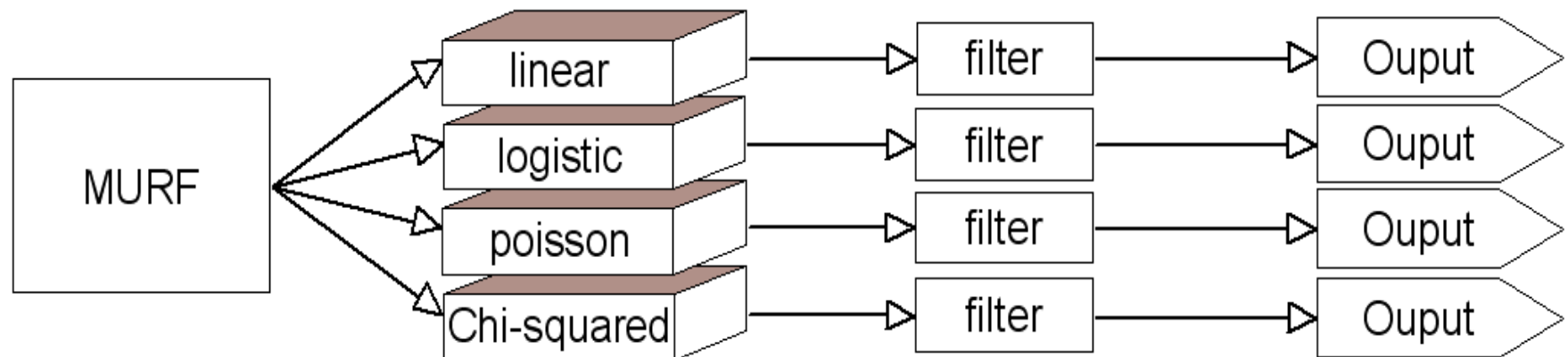
ABS research environment



Current researcher environment



Proposed future environment



Challenge:

- The ABS won't be able to develop tailored confidentiality methods/filters for every type of analysis currently used by researchers.
- The ABS proposes to identify the core requirements and incrementally develop this functionality over the next 2 years while exploring other access arrangements to facilitate more complex analyses (such as enhancing the existing ABSDL).



Questions?

