stralian
reau of
atistics

Data Collection and Processing – The Australian Experience

UN Economic Census Workshop, China 2005

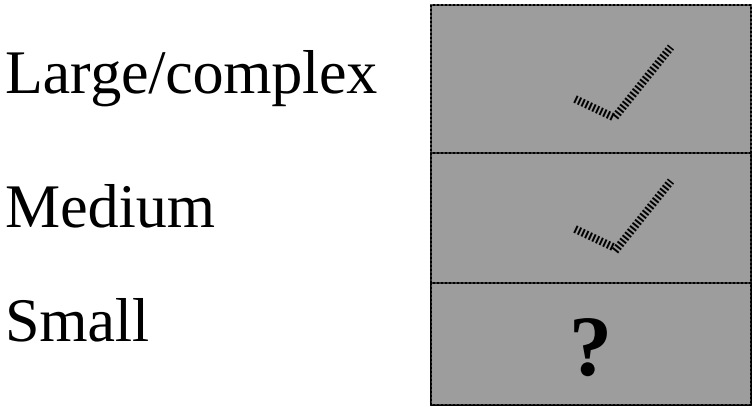
Presented by Eden Brinkley

Overview of presentation

- Aspects of data collection and processing for Annual Integrated Collection
  - Mode of collection
  - Questionnaire
  - Sample design
  - Estimation
- Integrated systems infrastructure
- Key points

Mode of collection

- Direct collection (about 20,000 businesses for core)



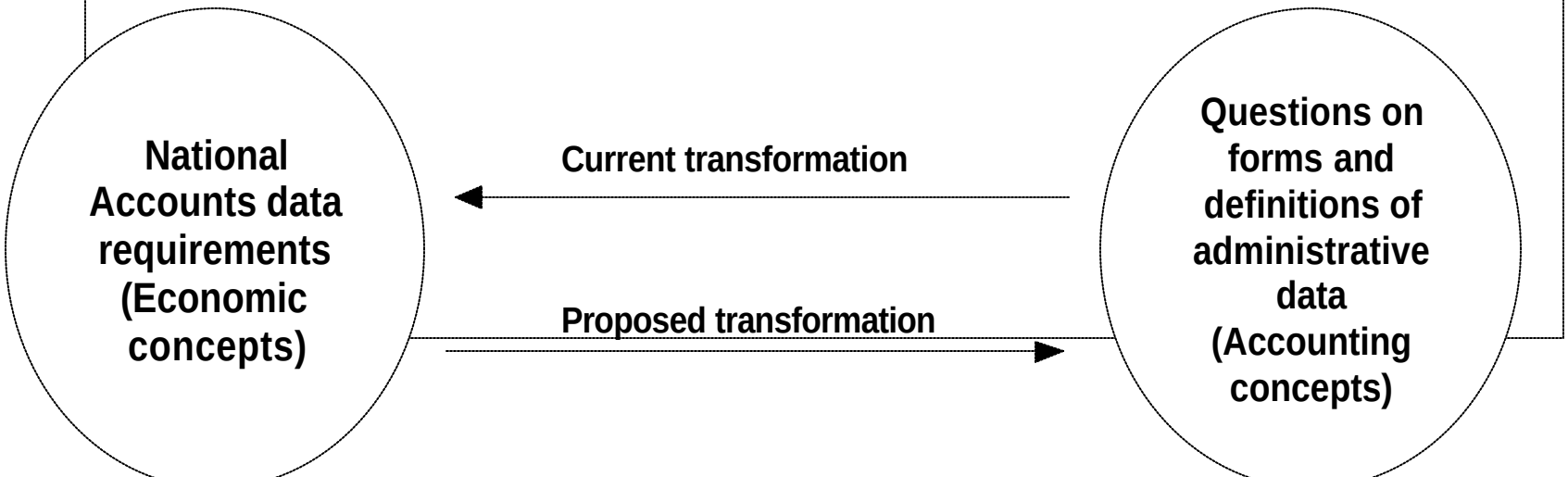
- Australian Taxation Office administrative data
  - annual Business Income Tax data
- Other administrative sources for government data, finance sector data and some agricultural data

Mode of collection (cont'd)

- In the last few years ABS has centralized its data collection activity into the Economic Statistics Data Centre
  - Administrative Data Acquisition Unit
  - Data Capture Unit
  - Provider Contact Unit
  - Business Register Unit
- Business Statistics Centres also created to process data (including editing) and deliver quality outputs

Mapping economic and accounting concepts

- Key National Accounts requirements specified in economic terms (e.g. gross output)
- Currently trying to map economic concepts to accounting concepts and questions on forms



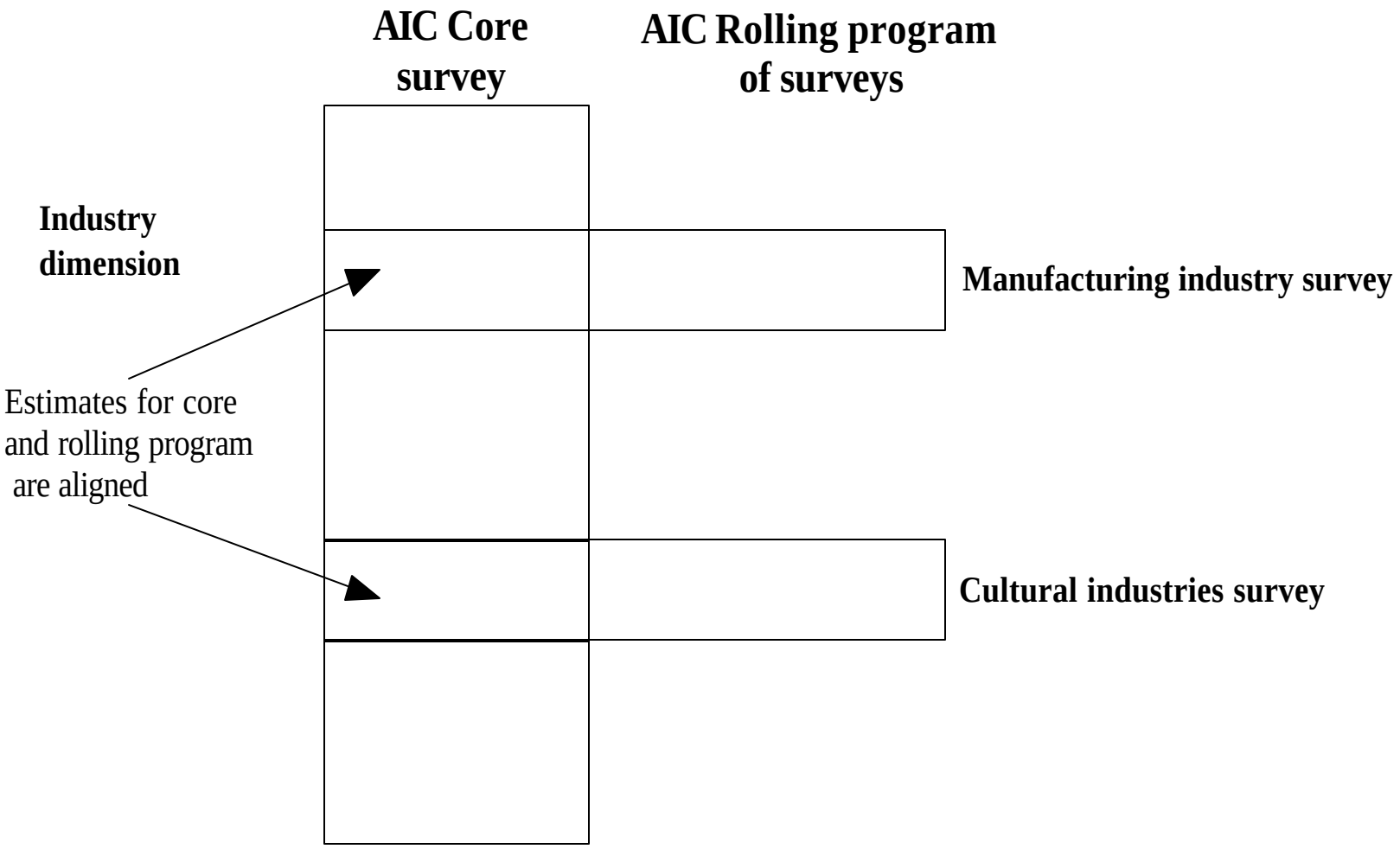
Questionnaire design

- Not decided on the future strategy for the AIC yet, but some rationalization of forms is required
- Investigate cost/benefit of having:
  - A core form and range of tailored supplementary forms for groups of similar businesses, versus
  - One large tailored form for groups of similar businesses
- Need to get a balance between tailored industry forms and generic forms

Integrated sample design

- Designing the sample such that industry estimates from the core and rolling program are aligned
- Sample for rolling program is effectively a top-up of core sample for the relevant industry

Alignment of estimates from AIC core and rolling programs



Sample design

- Stratification
  - Industry x sector x broad region x size
- Design for reliable estimates of movement as a first priority
- Design variables
  - industry value added, compensation of employees, gross operating surplus, gross fixed capital formation

Estimation

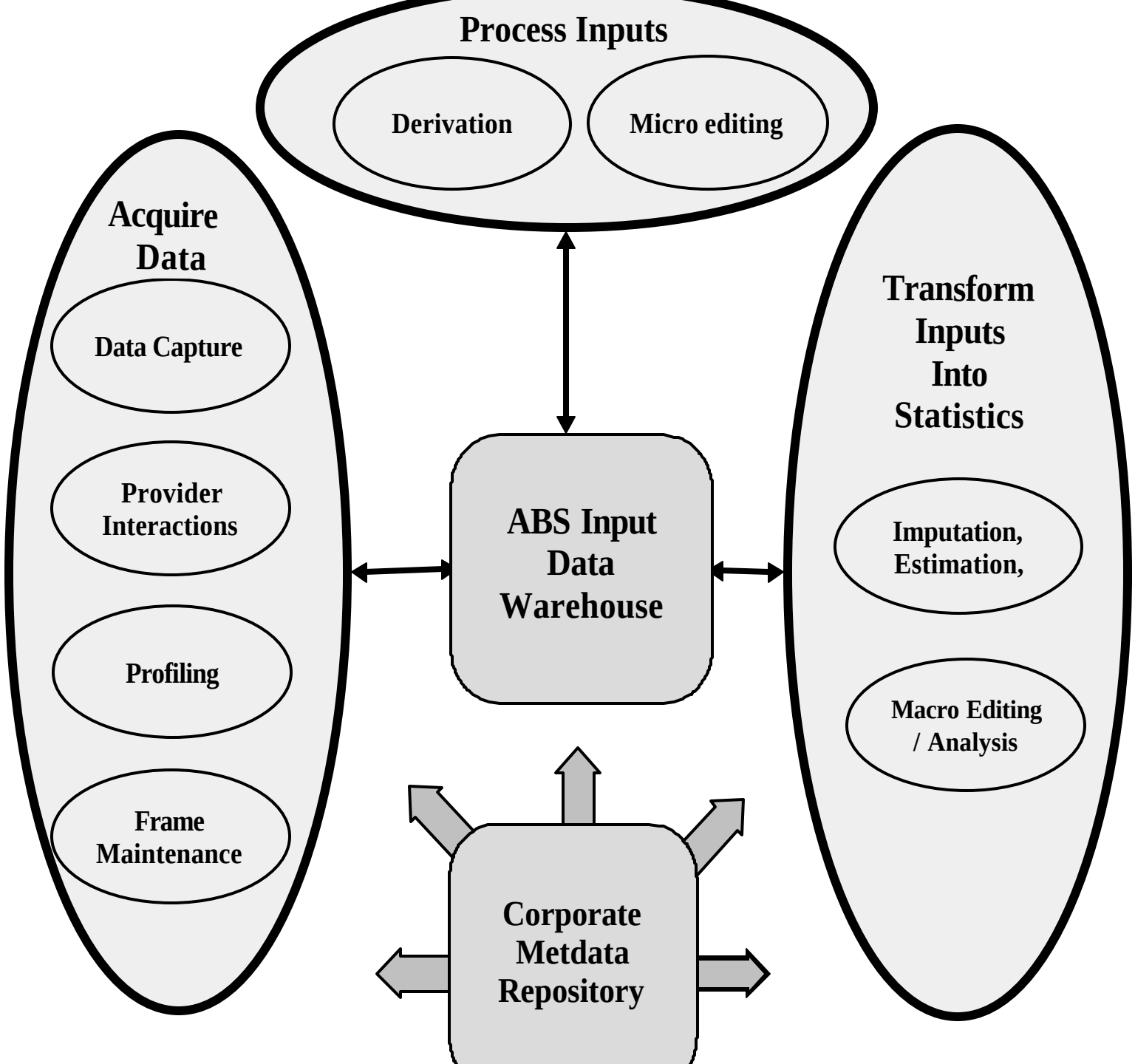
- Moving to using generalised regression estimation for most outputs
- Inclusion of non-employing businesses in scope leads to challenges in modelling some data items not found on the tax file (e.g. employment)
- Exploring options to release more timely data (no later than 12 months after reference period)

Integrated infrastructure

- ABS has already built the basis of an integrated environment
- Main components include:
  - Input Data Warehouse
  - Intelligent forms scanning facilities
  - Blaise for data capture and editing
  - Standard estimation and imputation system
  - Provider Integrated Management System
  - Central repository for storing aggregate data
  - Central metadata repository

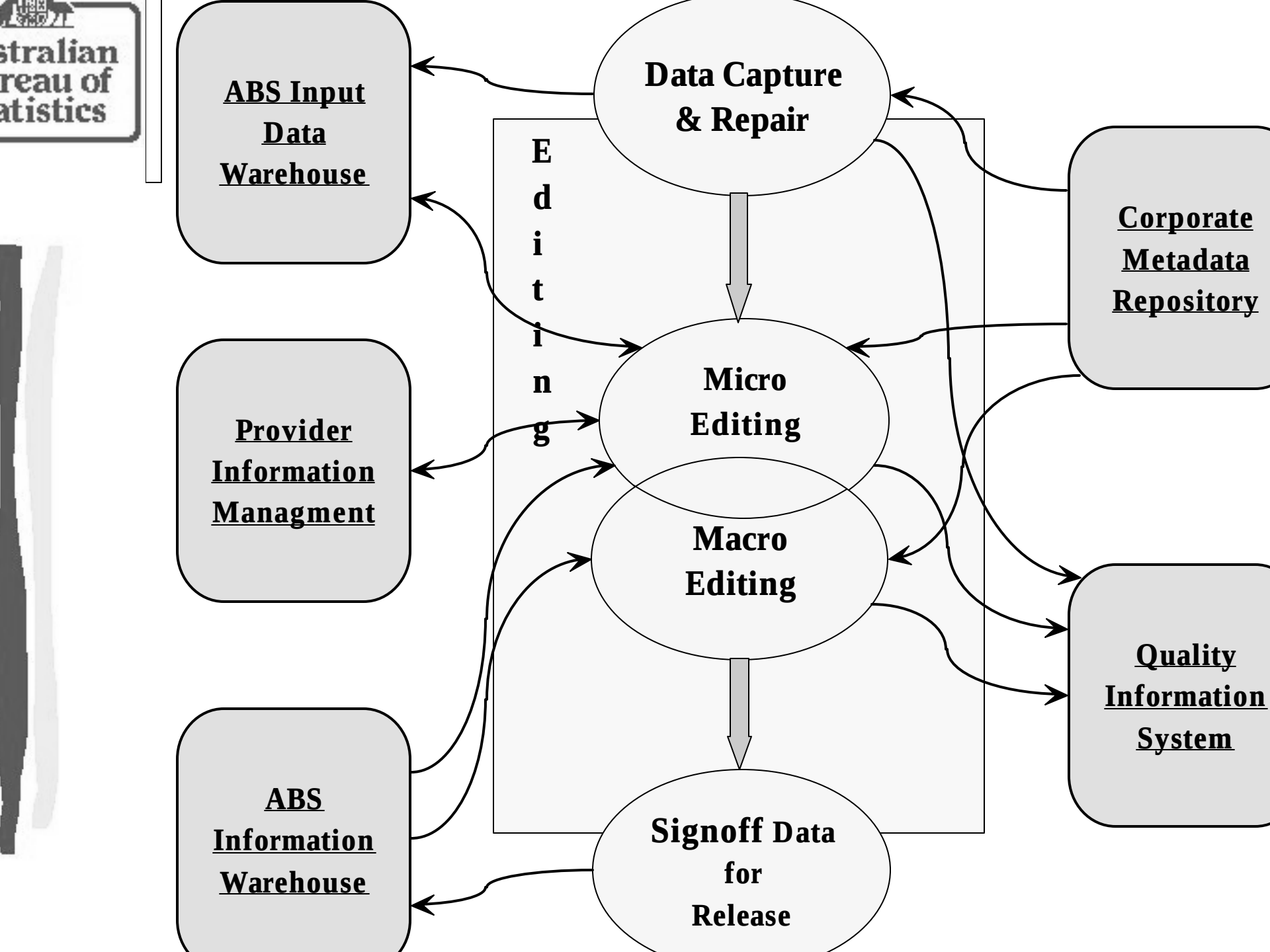
Input Data Warehouse

- Managed unit record data store
- Key enabler for re-engineering business processes. Expected to facilitate:
  - Enhanced statistical and analytical use of unit record data
  - Expanded use of administrative data
  - Research and implementation of improved practices and data management functions


Process Inputs
Derivation
Micro editing
Acquire Data
Data Capture
Provider Interactions
Profiling
Frame Maintenance
ABS Input Data Warehouse
Transform Inputs Into Statistics
Imputation, Estimation,
Macro Editing / Analysis
Corporate Metdata Repository

Editing re-engineering

- Aiming to deliver ‘fit for purpose’ data at minimum cost in appropriate timeframe
- Develop and implement best practice editing methodologies and systems
  - Developing a Data Editing Guide and training
  - Developing a detailed editing strategy for each dataset, integrated for the annuals
  - Addition tools for editing toolkit (e.g. significance editing and graphical drill down tools)


ABS Input Data Warehouse
Data Capture & Repair
Corporate Metadata Repository
Editing
Micro Editing
Provider Information Managment
Macro Editing
Quality Information System
ABS Information Warehouse
Signoff Data for Release

Key learning points

- Focus on priority client needs at each stage of the collection design and conduct
- Maximize use of administrative data for design, estimation, supplementation, etc
- Design outputs to ensure they are ‘fit for purpose’

Key learning points (cont'd)

- Integrated processes and systems should lead to improved efficiency and data quality
- Build systems from perspective of end user