

# Use of Administrative Data and Statistical Business Register

Valentin Todorov  
UNIDO

[v.todorov@unido.org](mailto:v.todorov@unido.org)

---

Regional Workshop on Industrial Statistics  
for the Arab Countries  
15-16 November 2011, Beirut

## Outline

- Introduction: definitions
- Benefits and issues of using administrative data
- Business Register
  - Functions
  - Data Sources
  - Data Structure
  - Record linkage/matching
  - Updating system
- GSBPM for maintenance of Business Register

## What are *Administrative Data*

**Definition 1:** Data available in records of government bodies administering various governmental programmes, mainly collected for the purpose of carrying out various non-statistical programs

**Definition 2:** A wider definition would include also organizations operating in private sector

**Examples:** tax register, social security database, register of companies (industrial register), register of Economic Chamber

## Benefits of *Administrative Data*

**Cost:** Surveys and censuses are expensive while administrative data can be cheap, often “free”

**Response Burden:** Reduced burden on data suppliers; Statistics can be compiled more frequently with no extra burden

**Coverage:**

- Full coverage of target population
- No survey errors and lower non-response
- Better small-area data

**Timeliness:** can be improved for some types of data but not for all

## Issues of *Administrative Data*

- Administrative units do not always coincide with statistical units
- Necessary to perform data parsing and conversion
- Different definitions and classifications: Conversion tables needed for different classifications
- Administrative and statistical priorities are often different
- Timeliness
  - Data may arrive too late
  - Data may relate to a different reference period
- Data from multiple sources: matching/linking issues, data conflicts

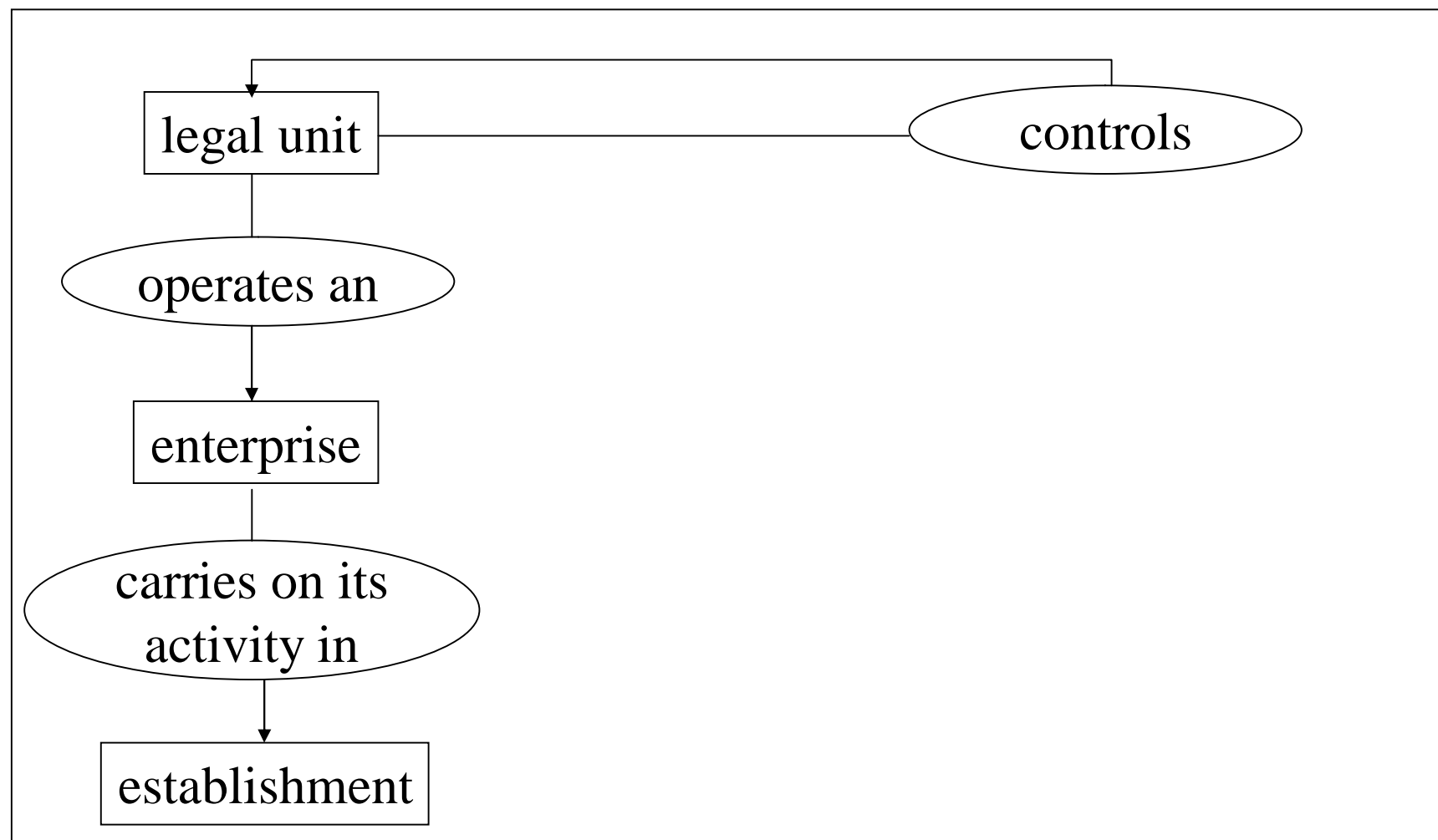
## *Business Register*

- Business Register is a list (computer file, database) of the enterprises and/or establishments in a country, with an identification number for each unit
- Business Register is an essential tool for industrial statistics:
  - Serves as a statistical frame for industrial surveys
  - Serves as a basis for grossing-up results from sample surveys to produce business population estimates
  - Serves as an important tool for managing an industrial census
  - A system for updating the BR is essential for the preparation of reliable time series for employment, value added and business demography in the industrial sector

## *Recommendations (IRIS 2008)*

- Reliance on a comprehensive BR is recommended (previously a separate industrial directory was recommended)
- Statistical units: both establishments and enterprises with links to each other
- Strategy for data collection: the importance of using administrative data is stressed
- Practical need for a cut-off: coverage of units of more than a certain size (usually employment size)

## *Conceptual model*





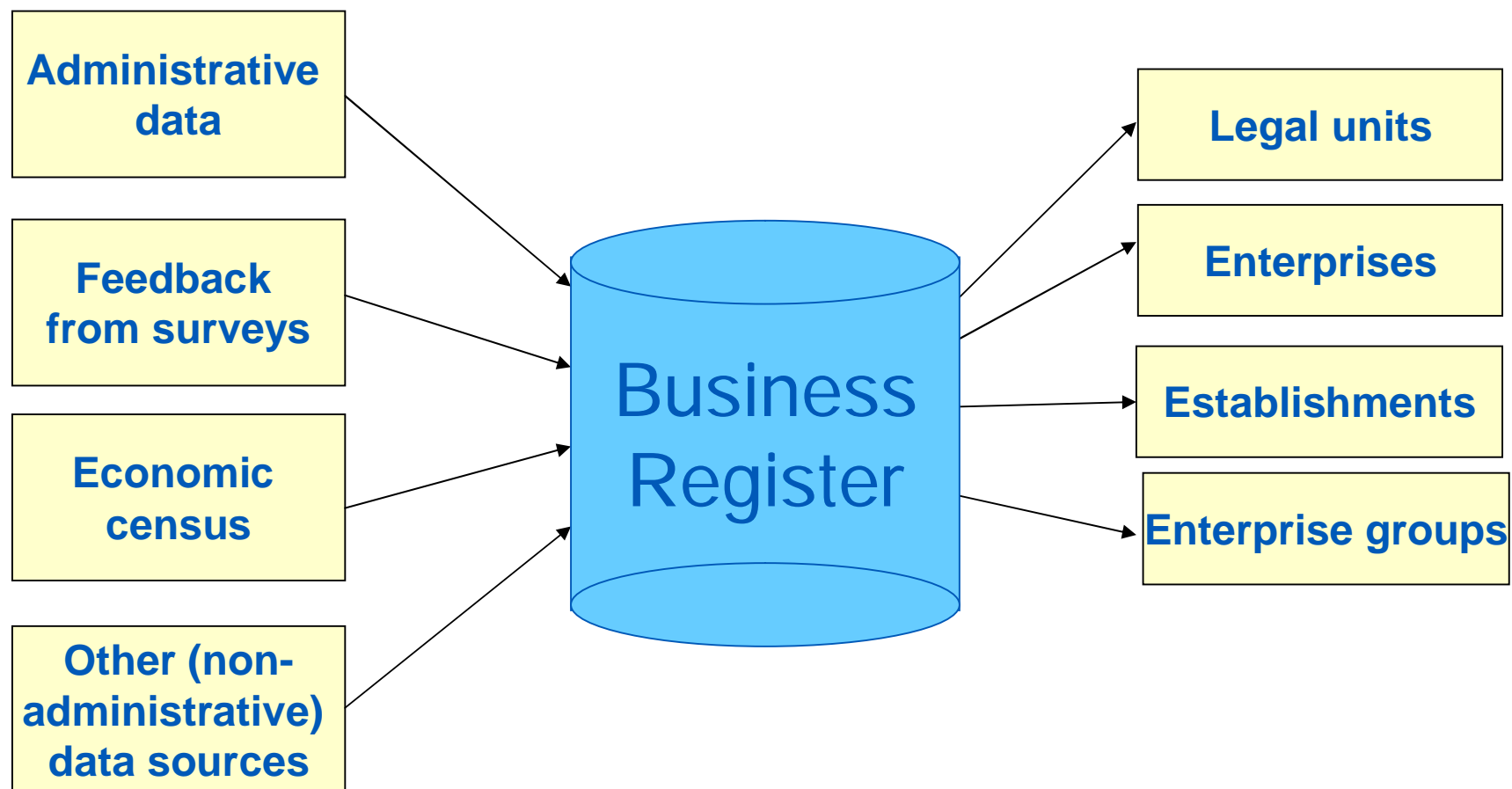
## *Functions of BR: (i) Traditional*

- Frame for industrial surveys: provide appropriate ID and address data. The results of the survey are important input for updating the register. The register is used as a basis for grossing-up results from sample surveys to produce business population estimates
- Survey support: BR stores important contact data to facilitate the filling in of the questionnaires; Can document the workflow of the conduction of the survey
- Comprehensive survey documentation: which units where selected for which surveys; basis for response burden calculations, can be enlarged with information related to response burden

## *Functions of BR: (ii) Emerging*

- Documentation of classification codes: needs empirical and up-to-date information on the production structure of the enterprise; indication of the relative importance of the activities
- Cooperation with other statistical institution (e.g. the National Bank)
- Unique statistical database
- Other statistical uses: e.g. business demography statistics
- Link to administrative data: For example updating turnover data from tax register; the link to the social security database for updating the number of employees, etc.
- Other non-statistical uses: depends on the legal basis

## *Data Sources of BR*



## *Data Sources of BR*

- Tax Registers: can serve for both entries and exits
- Single administrative source
- Multiple administrative sources
  - Lack a common identifier (matching necessary)
  - Duplicate entries (due to matching errors)
  - Complete coverage cannot be expected
  - No source may be available to document closures (if tax register is absent)
  - Costly field checks needed
- Economic Census
- Feedback from industrial surveys (very effective for exits)

## *Data Structure of BR*

1. Name and physical location;
2. Mailing address;
3. Name and address of the headquarter
4. Kind of economic activity, description or code;
5. Legal organization - incorporated or unincorporated;
6. Type of ownership: public (by central, state and local governments); national, private and foreign controlled;
7. Number of persons employed;
8. Volume of sales or value of output;
9. Source and date of information

## Data Structure of BR

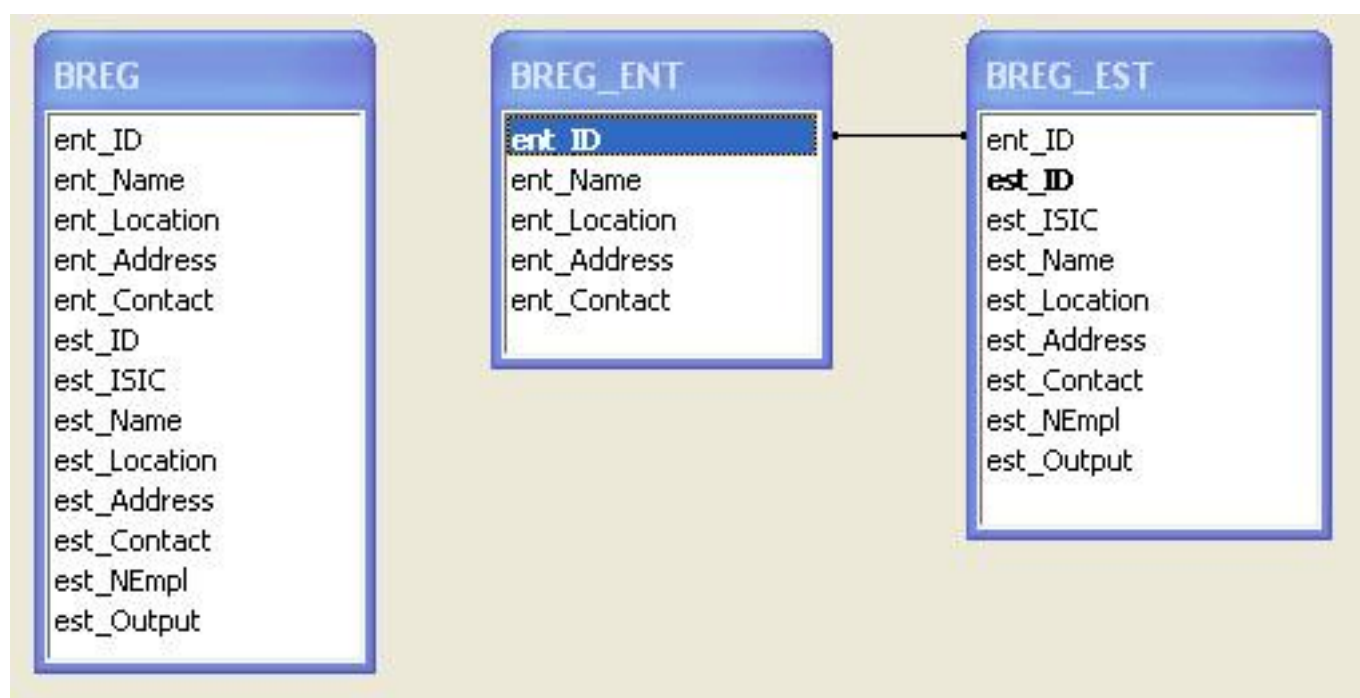
■ BREG : Table													
	ent_ID	ent_Name	ent_Location	ent_Address	ent_Contact	est_ID	est_ISIC	est_Name	est_Location	est_Address	est_Contact	est_NEmpl	est_Output
	0001	AAA				00011	1511	AAA-1				0	0
	0001	AAA				00012	1511	AAA-2				0	0
⌵	0001	AAA				00013	1511	AAA-3				0	0
	0002	BBB				00021	1600	BBB-0				0	0
	0003	CCC				00031	2301	CCC-1				0	0
*												0	0

BREG_ENT : Table						
		ent_ID	ent_Name	ent_Location	ent_Address	ent_Contact
▶	+	0001	AAA			
	+	0002	BBB			
	+	0003	CCC			
*						

BREG_EST : Table									
	ent_ID	est_ID	est_ISIC	est_Name	est_Location	est_Address	est_Contact	est_NEmpl	est_Output
▶	0001	00011	1511	AAA-1				0	0
	0001	00012	1511	AAA-2				0	0
	0001	00014	1600	AAA-3				0	0
	0002	00021	2301	BBB-1				0	0
	0003	00031	2301	CCC-1				0	0
*								0	0

## *Data Structure of BR*



## *Record linkage and matching*

- When multiple data sources are used the system will require extensive linkage and matching
- Two kinds of duplication:
  - Records which are already in the register
  - Introduction of duplicates into the register
- Easy if there is a common reference number for businesses (*Exact Matching*)
- Otherwise apply *Probabilistic Matching* - determining the probability that records from different sources should match, using a combination of variables
- Perform quality checks
- Appropriate use of existing software should be made



## *Recoding (ISIC R3 → ISIC R4)*

- In principle, the relationship between elements in R3 and R4 may be one of the following four types:
  - **1:1 cases:** One element in the old classification can be found in the new classification in the same way.
  - **1:n cases:** One element in the old classification has been subdivided into two or more elements in the new classification.
  - **n:1 cases:** Several elements in the old classification have been aggregated to form one element in the new classification.
  - **m:n cases:** Parts of different elements of the old classification have been reassembled so that the new elements are not simply aggregates of old elements or disaggregates of one old element. These are typical configurations in classification revisions.
- It is clear that only 1:1 and n:1 cases can be recoded automatically.

## *Updating system*

- It is preferable to utilize an integrated register application with the following basic modules:
  - Importing of external sources
  - Parsing and editing data from external sources
  - Matching (sources against each other and against register)
  - Entering of field data, validation and storage
  - Managing the register (edit, ad hoc edition, etc.)
  - Querying and reporting
  - Managing of the annual industrial survey and other surveys

## Business Register and GSBPM

- Generic Statistical Business Process Model (GSBPM)
  - Defined by the Joint UNECE/Eurostat/OECD Work Sessions on Statistical Metadata (METIS)
  - Should be seen as a flexible tool to describe and define the set of business processes needed to produce official statistics.
  - To define and describe statistical processes in a coherent way
  - To standardize process terminology
  - To compare and benchmark processes within and between organisations
  - To identify synergies between processes
  - To inform decisions on systems architectures and organisation of resources

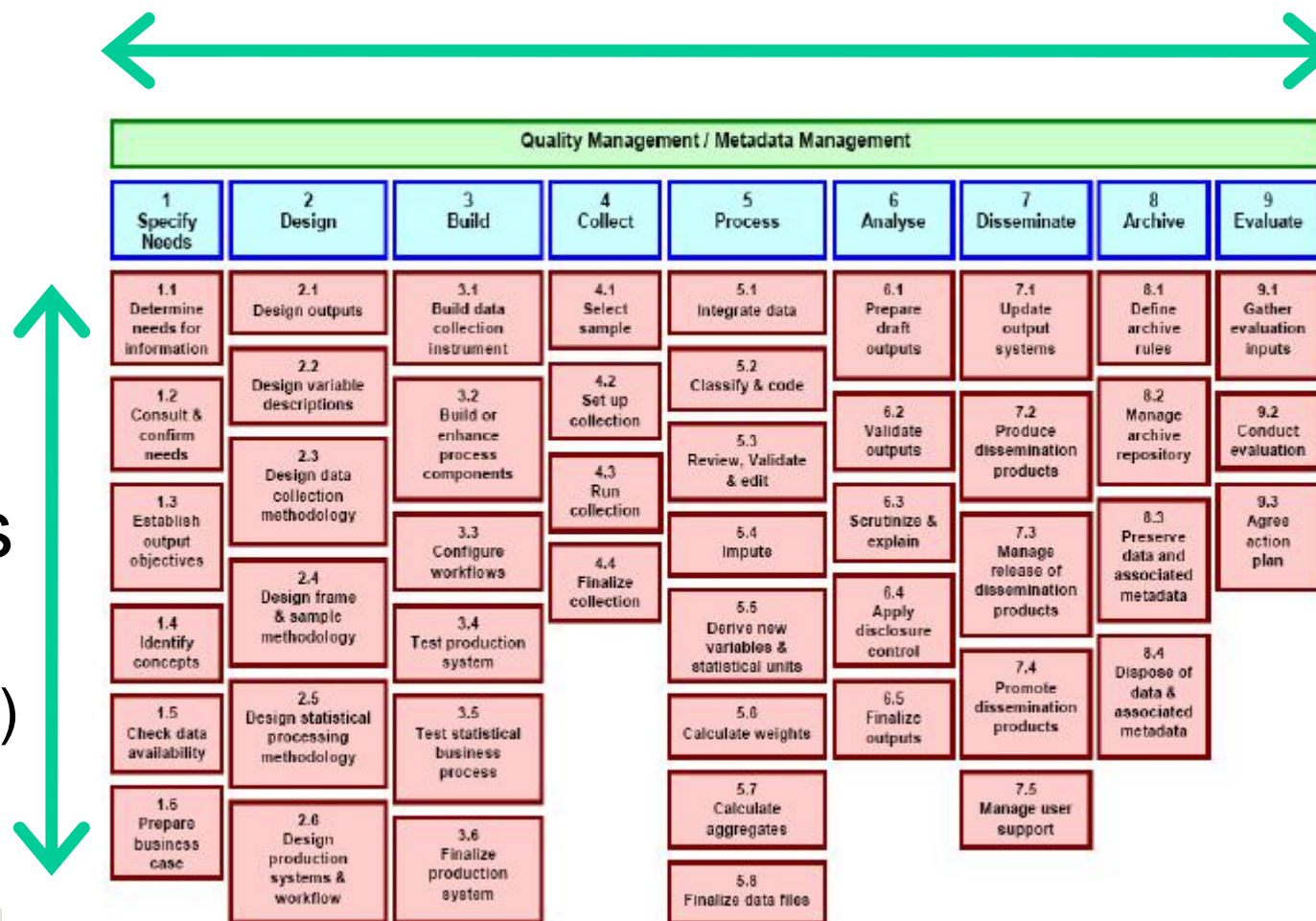
## Generic Statistical Business Process Model

Process

Phases

Sub-  
processes

(Descriptions)



Source: METIS (2009)

## Application of GSBPM to Business Register

- Business register maintenance is a continuous activity and not a single process
- But Business registers operate on:
  - Inputs “collected” from different sources
  - A sequence of processing and analysis
  - Outputs – statistics and sampling frames
- Therefore BR maintenance can be seen as similar to other statistical production
- GSBPM can be applied to statistical business register maintenance: clear potential benefits in terms of shared knowledge, methods and tools

## Summary and conclusions

- There are many benefits in using administrative data; the main benefit is usually reduced costs. There are also problems but for many of them exist suitable solutions
- Business Register is an essential tool for industrial statistics
- The effects using business register and linkages to administrative data are: increased data quality; increased efficiency; reduced costs; reduced response burden
- GSBPM could be applied to statistical business register maintenance