

THE JUNCTION BETWEEN EXTERNAL DATA AND STATISTICS DATA

**Is it possible to optimise the roles of data
suppliers and users?**

The UNECE Working Session in Madrid, October 2003

**Topic iv) - Data Editing by Respondents and Data
Suppliers**

Seppo Laaksonen

and

Ismo Teikari

Statistics Finland, Finland

Supported by

Pekka Myrskylä (Statistics Finland), Mikko Pellinen (Finnish Centre for Pensions) and
Kalevi Heiliö (Central Taxation Agency)

Targets



Long term

Targets

- Use as much administrative and other existing data as possible
 - External and Internal
- Try to manage the work so that the data would be edited at as low/early level as possible
- Make the final editing yourself and give feedback to previous editors
- Exploit the data completely

Primary Units

Examples



- **People in a country**
- **Workers in enterprises**
- **Tax payers (people or enterprises)**
- **Insurance payers (people)**

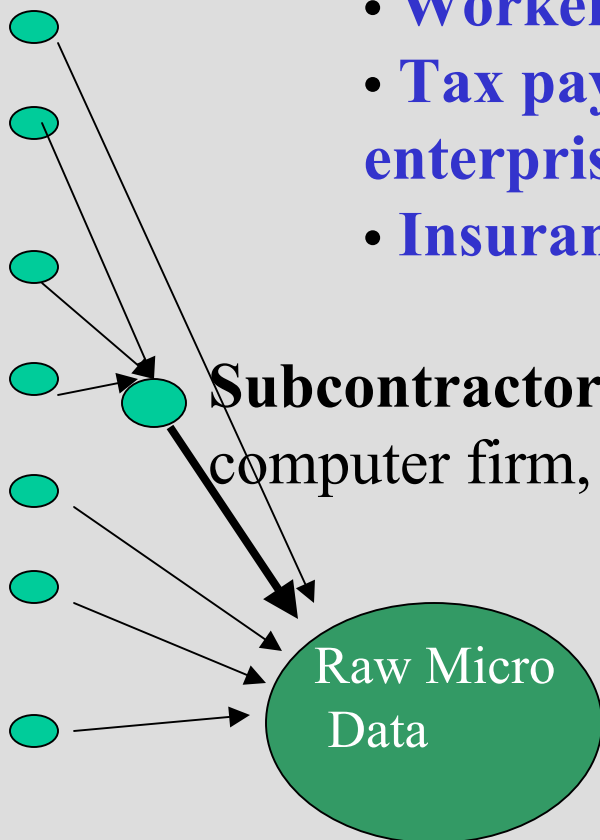


Illustration of the Roles of Data Suppliers and Statisticians A

Primary Units

Administrative Supplier

- People in a country
- Workers in enterprises
- Tax payers (people or enterprises)
- Insurance payers (people)



Subcontractor possible:
computer firm, local authority

Central Population register
Taxation Agency
Finnish Centre for Pensions (ETK)
Employers' organisations

Illustration of the Roles of Data Suppliers and Statisticians B1

**Primary
Units**

**Administrative
Supplier**

Editing Principles

Taxation Agency:

- (i) The data arrives on time and are complete and correct, including the required format of the data.
- (ii) The data arrives late but are complete and correct.
- (iii) The data are not correct.
- (iv) The format of the data are not correct.
- (v) The form of the data are correct but do not correspond the truth
- (vi) There are logical discrepancies in the data.
- (vii) The data are correct and arrived on time but do not correspond to the truth from the point of view of taxation rules (abuse).

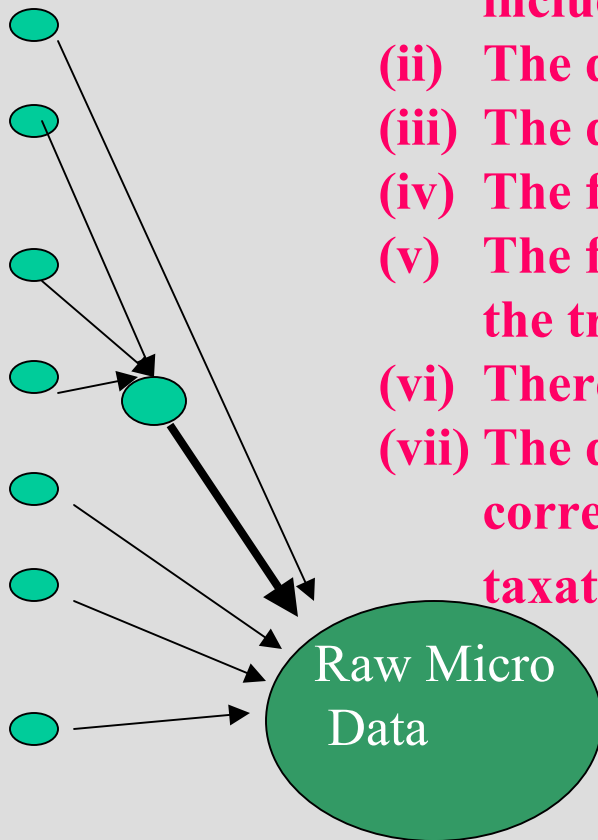
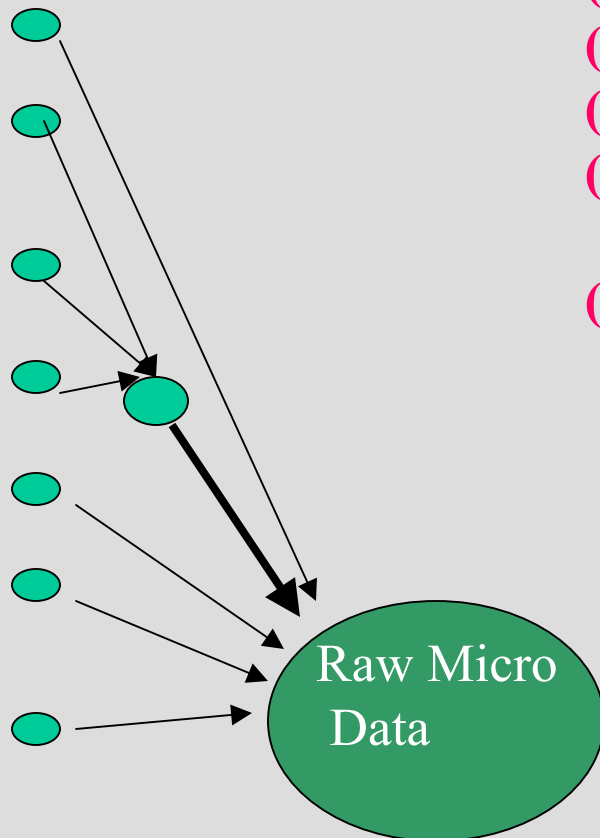


Illustration of the Roles of Data Suppliers and Statisticians B2

**Primary
Units**

**Administrative
Supplier**

Editing Principles



Finnish Centre for Pensions (ETK):

- (i) The form of message has been checked.**
- (ii) The logic of the message within each message.**
- (iii) Checked new information against the old one.**
- (iv) Searching for problematic cases using a computer program made for this purpose.**
- (v) Comparing the values with the values of pension institutes (coherence).**

Every month has been taken an aggregate output (tables) and these have been compared with previous outputs. The outputs have been used as such but also these are useful for checking results at macro level.

Illustration of the Roles of Data Suppliers and Statisticians B3

**Primary
Units**

**Administrative
Supplier**

Differences

Employers' organisations

- The business units are not similar in all cases as in the BR (member enterprises vs. BR units).
- Some statistical classifications are not of high interest to the TT&PT.
- The TT&PT have started to use some Statistics Finland classifications, especially our register of formal education. This has harmonised results for recent years, but in a long term, there are problems.
- Occupations are 'real-life'-based in all wages statistics (quite different) but can be harmonised well by converting to standardised ISCO classification.

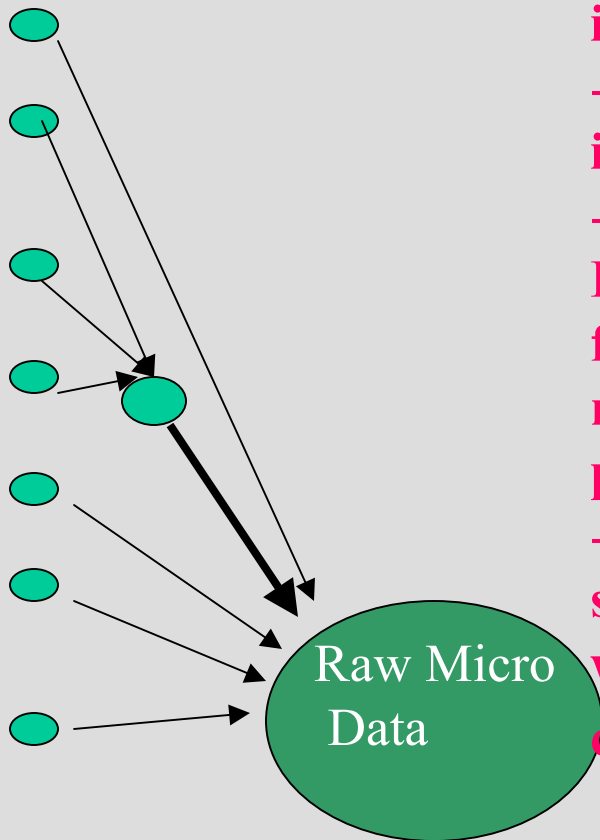


Illustration of the Roles of Data Suppliers and Statisticians B4

**Primary
Units**

**Administrative
Supplier**

Differences

Employers' organisations (continued)

- Some variables are difficult to measure, e.g. weekend and other special work time, overtime work and over time wage, bonuses and tenure, but less motivated for the TT&PT than for Statistics Finland. This has sometimes led to impute some items to TT&PT data.
- Part-time and temporary workers may be more often missing from the TT&PT data than from Statistics Finland data.
- The TT&PT data do not cover all businesses of the sectors desired to examine, esp. small businesses. Some additional data collection has been made in Statistics Finland, the purpose being to fill these gaps. This makes the statistics production system quite complicated, giving challenging jobs for statistical methodologists.

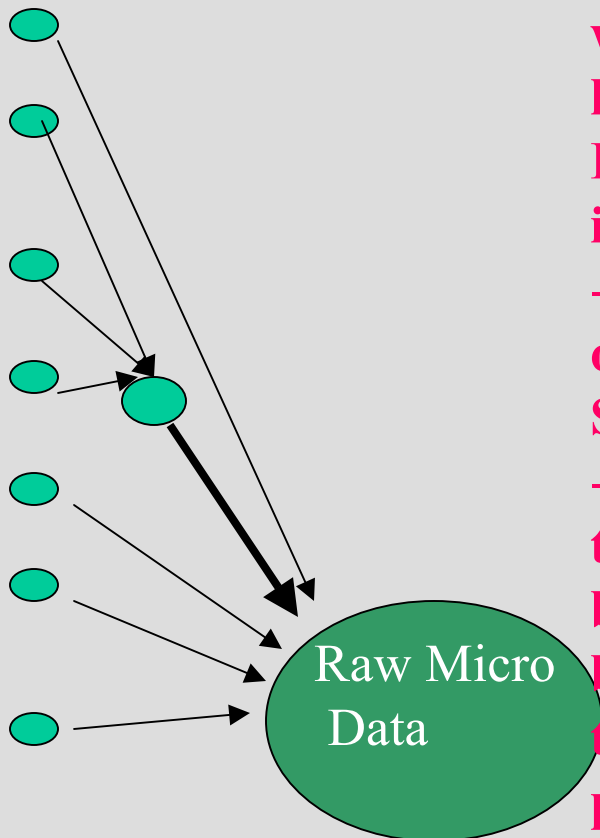
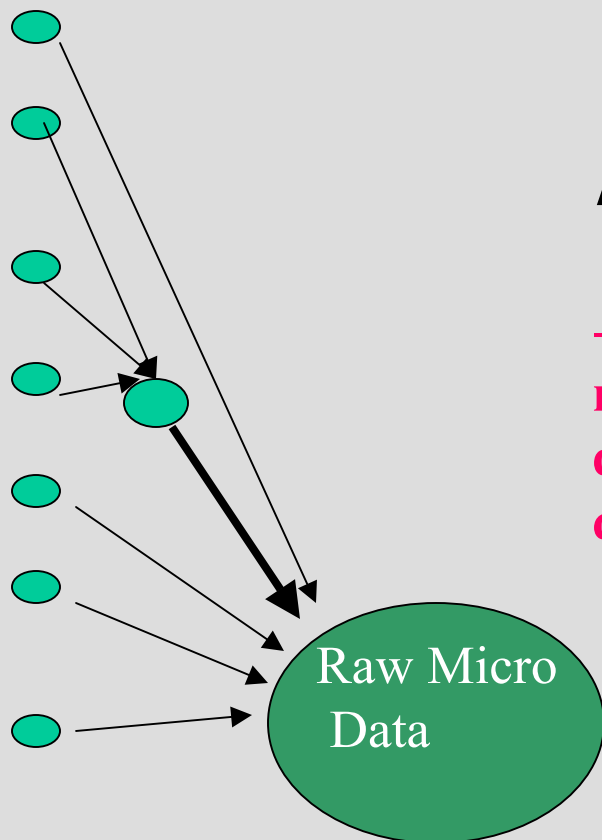


Illustration of the Roles of Data Suppliers and Statisticians B4

**Primary
Units**

**Administrative
Supplier**

Differences



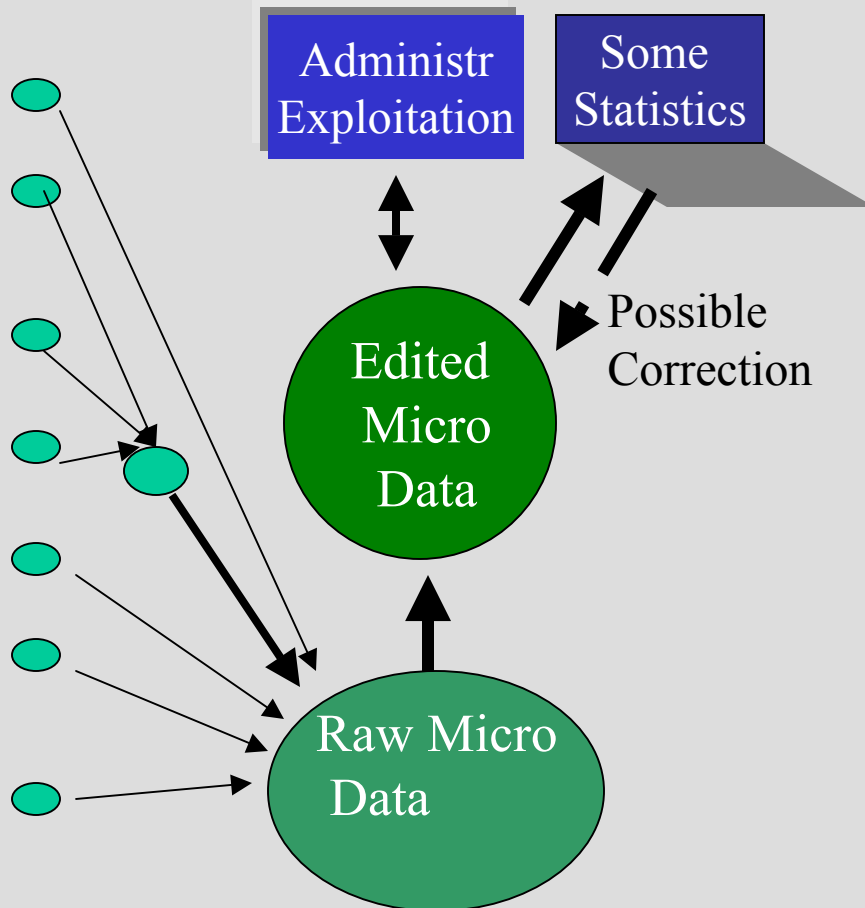
Employers' organisations (continued)

-For econometric research the TT&PT data are much more used than any Statistics Finland data due to their long tradition with about similar quality (longitudinal analysis will be reliable)..

Illustration of the Roles of Data Suppliers and Statisticians B5

**Primary
Units**

**Administrative
Supplier**



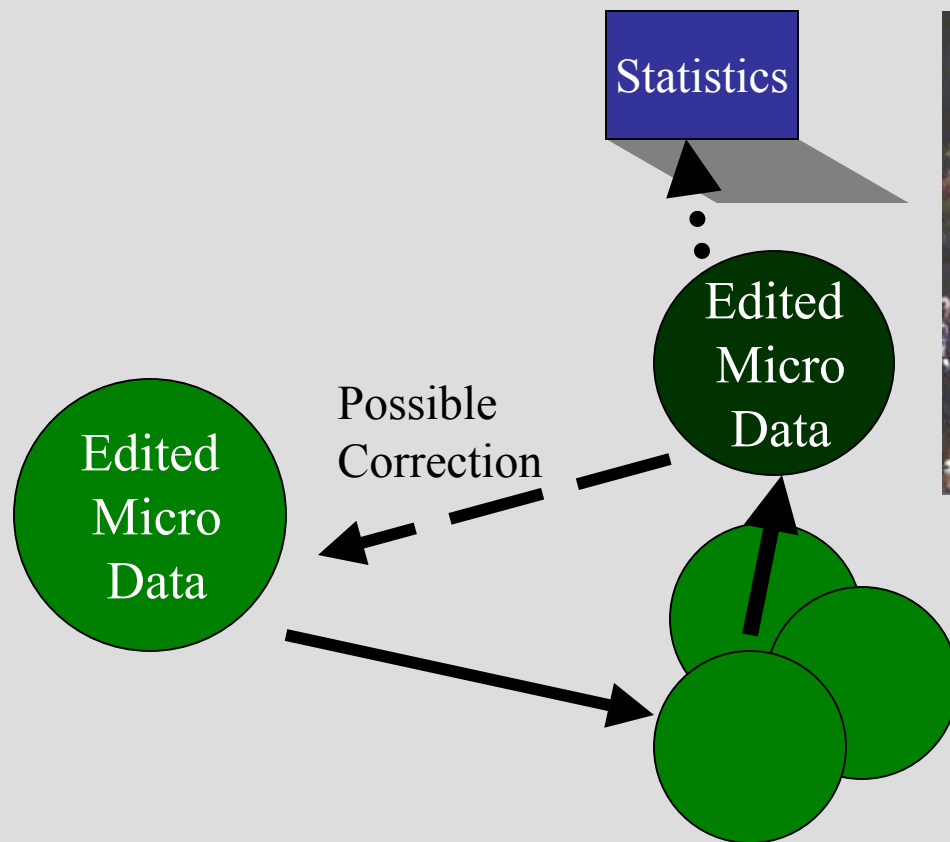
Most authorities publish quite simple aggregated statistics only, but the employers' federations some more demanding figures especially for wage bargaining

Illustration of the Roles of Data Suppliers and Statisticians C

**Primary
Units**

**Administrative
Supplier**

Statistical Agency



Aim= To Use as such
or after Linking with
Other Data
(Survey, Register)

Illustration of the Roles of Data Suppliers and Statisticians D

Example *Linked Longitudinal Employment Data (LLED)*

Key Questions for Editing

- (i) Who is a resident and where this resident is living (to include the co-ordinates of the building is the target),
- (ii) What is the employment status of a resident (incl. labour force status and if working occupation and other information),
- (iii) Where an employed resident is working (again the target is to get the co-ordinates of the building).



Statistical Agency

Aim= To Use as such
or after Linking with
Other Data
(Survey, Register)

Illustration of the Roles of Data Suppliers and Statisticians E1

Example *Linked Longitudinal Employment Data (LLED)*

Specific problems:

- The central population register does not indicate residents to business buildings (premises).
- LKAU'S in municipality and government, and also in private business entities are not well enough available in registers. It is better to organise an additional inquiry to these units using the information about the previous file and asking respondents to specify in which LKAU each person is working. (electronically and manually)

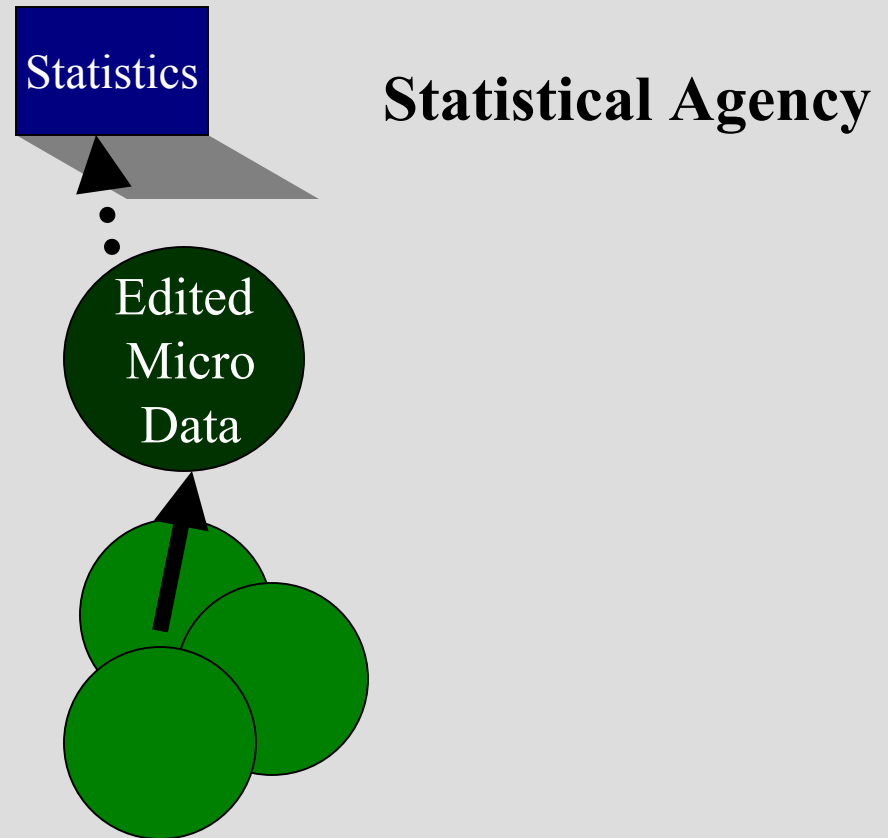


Illustration of the Roles of Data Suppliers and Statisticians E2

Example *Linked Longitudinal Employment Data (LLED)*

Specific problems:

- LKAU information is not good for all large businesses which have a high number of local activities (e.g. banks, kiosks, franchise firms).
- The same concerns firms which rent workers to other firms.
- A lot of difficulties in handling addresses
- A high number of simultaneous work-contracts for one person.
- The period of earning is not always correct.

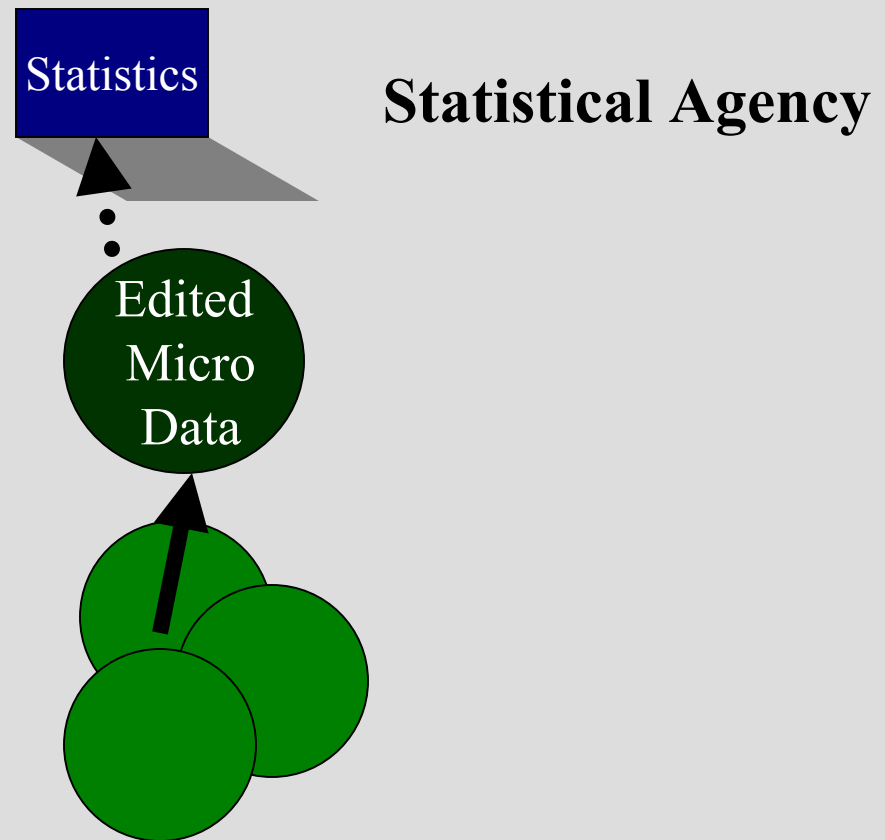


Illustration of the Roles of Data Suppliers and Statisticians E3

Example *Linked Longitudinal Employment Data (LLED)* Plus register-based Census

**Specific problem: How to
'impute' occupation:**

(•) The best sources were found from the yearly wage statistics which cover almost completely municipalities and government entities, and the members of the two employers' federations.

All these sources contain an occupation code for each employee in November or December of 2001. This code can be quite well converted to the ISCO standard.

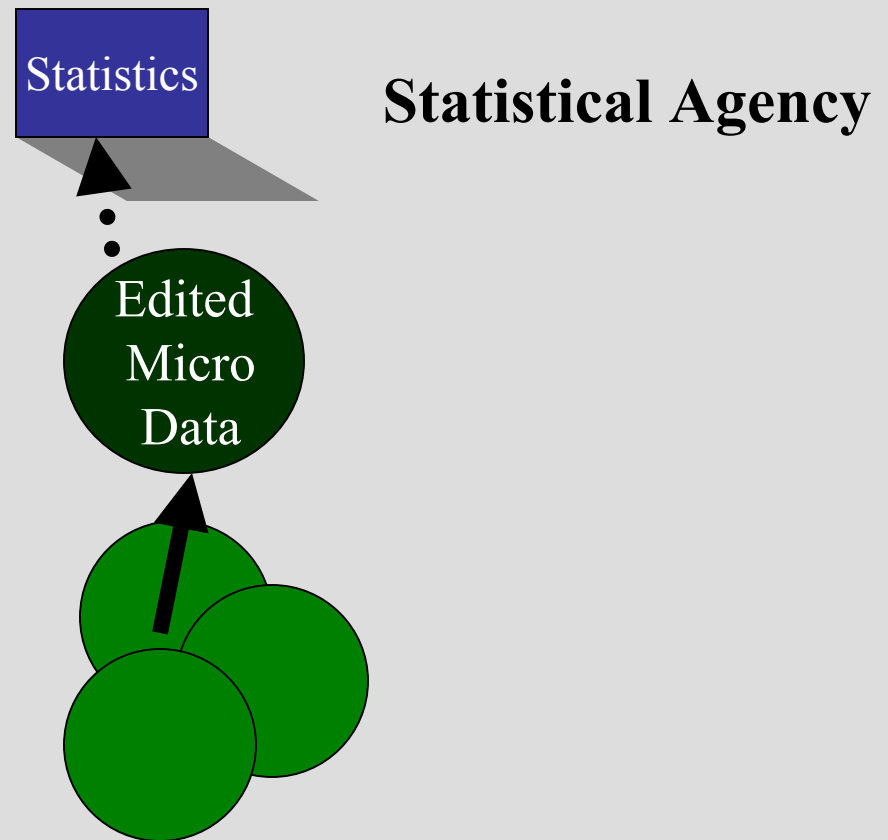


Illustration of the Roles of Data Suppliers and Statisticians E4

Example ***Linked Longitudinal Employment Data (LLED)***
Plus register-based Census

**Specific problem: How to
'impute' occupation:**

(••) The occupation of employed for smallest enterprises (less than 5 persons based on the BR) was imputed logically so that the same occupation had given to all persons deducing from the main activity of the enterprise (e.g. if the activity of an enterprise was hairdresser's then the occupation for all was hairdresser).

(•••) For other enterprises, a survey was conducted to indicate the occupation for each employee found from the employment register. This was the most expensive part of the 2001 census.

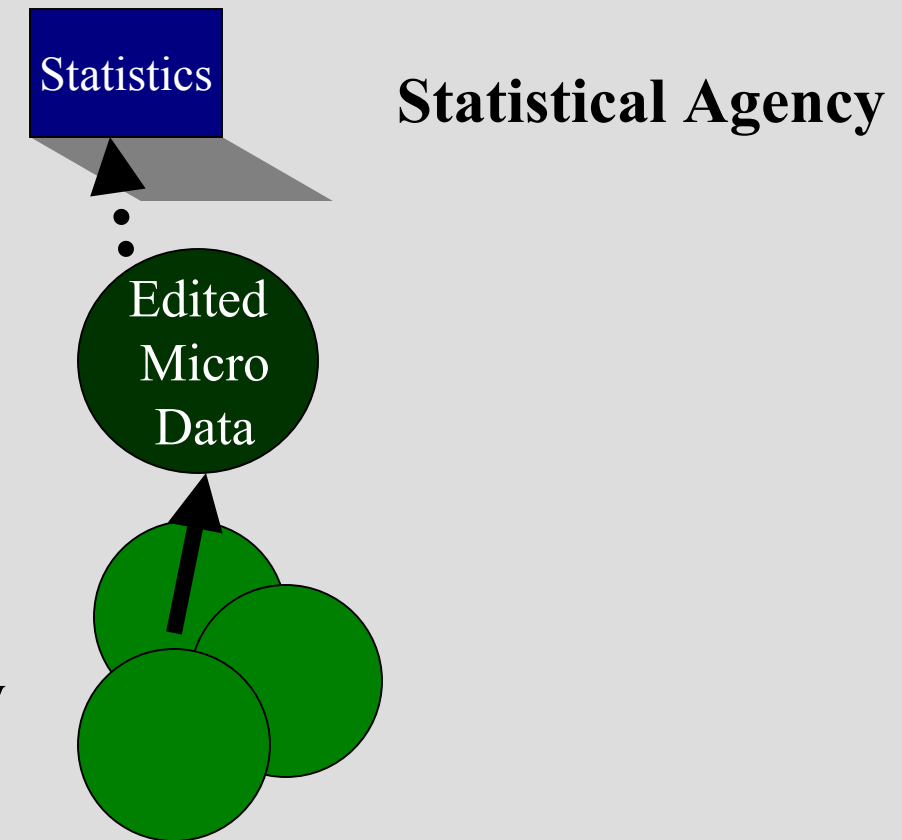


Illustration of the Roles of Data Suppliers and Statisticians E5

Finally, the role of **end-users** is very important in validating the data in a long term. The importance of researchers cannot be overrated. In particular, the researchers who are exploiting micro data (often based on several data sources and from a number of time periods) have been very useful for us. They are critical and ready to give feedback to statistics. It is not always possible to correct individual data a long time after the reference period, but these gaps may be marked, and the observed problems may be corrected in coming years.

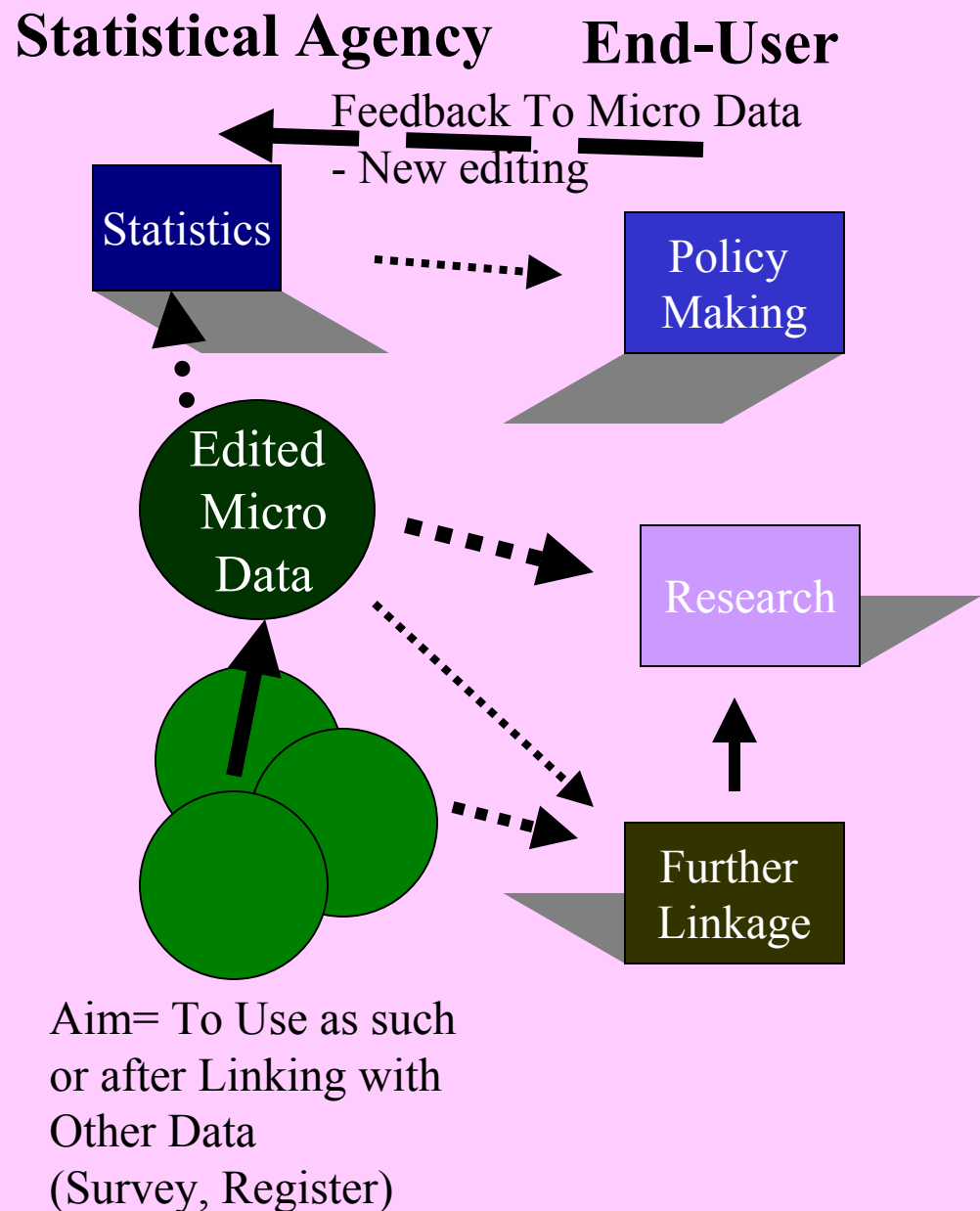


Illustration of the Roles of Data Suppliers and Statisticians F

Primary Units

Administrative Supplier

Statistical Agency

End-User

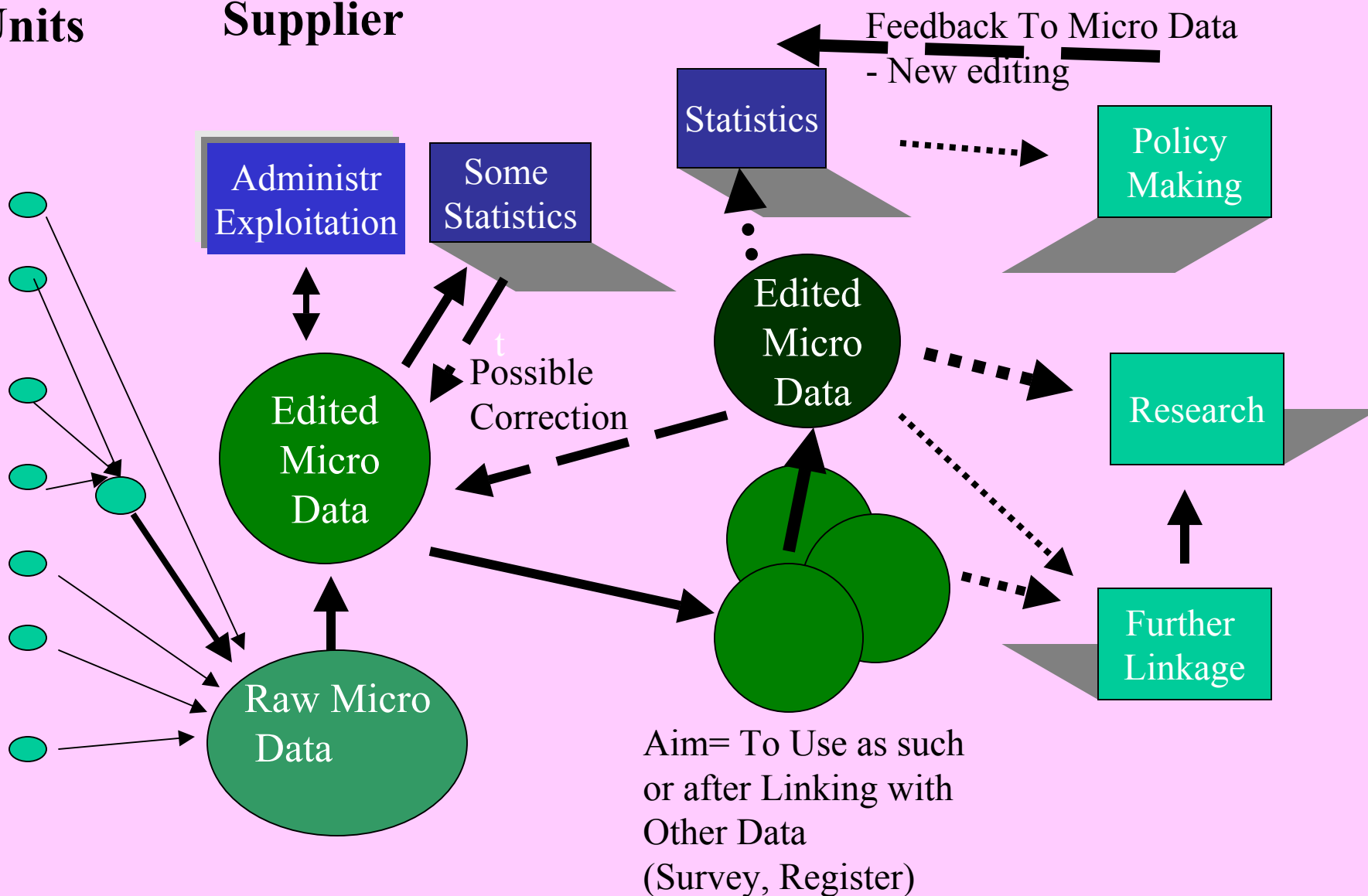


Illustration of the Roles of Data Suppliers and Statisticians G

Thank you for your attention
KIITOS

