

## **Mandate and Work Plan on Use of Administrative Data Sources for Energy Data Collection and Validation**

### **Use of Administrative Data Sources**

Administrative data are increasingly useful for government agencies all over the world as the current administrations continues to encourage data analytics and evidence based evaluations of programs.

The data are not collected for statistical purposes but for recordkeeping for example tracking participants, registrants, employees, regulate flows of goods such as fuels and peoples across borders, to administer benefits such as pensions or obligations such as taxations (for individuals or businesses). Administrative data are collected with a specific decision making purposes which makes every recorded units identity crucial compared to statistical records where no action concerning an individual or a business is intended or even allowed, the identity of individuals/businesses is of no interest once the database has been finalized.

Using administrative records presents several advantages to a statistical agency and to analysts. Demands for statistics on all aspects of our lives, our society and our economy, environment continue to grow. These demands often occur in a climate of tight budgetary constraints in the authorities. Statistical agencies also share with many respondents a growing concern over the mounting burden of response to surveys. Respondents may also react negatively if they feel they have already provided similar information (e.g. revenue) to administrative programs and surveys. Administrative records, because they already exist, do not incur additional cost for data collection nor do they impose a further burden on respondents. Advancements in technology have permitted statistical agencies to overcome many of the limitations caused by processing large datasets. For all these reasons, administrative records are being used increasingly for statistical purposes

Use of Administrative data is an activity that fulfils the responsibility of a fundamental principle of official statistics “official statistics authorities that meet the legislations from government to decrease response burden have to start to change the system of using more administrative data. The increase of use of administrative data are also correlated with more discussion about digitalization.

Statistical use of Administrative data includes:

- Use for Survey frames, directly as the frame or to supplement/update an existing frame (SAMU)
- Replacement of datacollection (use of taxation data for small businesses)
- Use in editing and imputation
- Direct tabulation
- Indirect use in estimation (auxiliary information in calibration estimation, benchmarking)
- Survey evaluation, including data confrontation (comparison between survey estimates and administrative data estimates)

### **Problem with use of Administrative data includes:**

- Lack of quality control over data
- Missing values (incomplete files)
- Different concepts (definitions) lead to bias
- Coverage problems
- Timeliness of data (without statistical control) wrong time
- Costly computer system for clean and control to make data useful (information security, handling data, storage data, document data, metadata)

### **Use of Administrative data policy/principle**

It is for all authorities those responsible for statistical area in Swedens statistical system and those who inspectorate guiding principle that whenever a new survey is conducted the authority should look after an administrative data source (record or observationregister from an already conducted survey) first before the authority starts conduct a new survey in their own purposes.

The problem in Sweden statistical system is the re-use of data between authorities because of the confidentiality of information which protects enterprises information in 70 years.

The vice versa example is when other authorities make their inspectorate with administrative data in the way that we can re-use the information in official statistics. That have been done in electricity price survey where SEA collects data directly from Energy market inspectortate (administrative record).

### **statistical confidentiality**

It is important to consider the scope of data which can be publically disseminated. The following should be considered.

- Consider privacy implications of the publication of information from administrative records. Sweden statistical Act provides all responsible authorities to access administrative data for their statistical purposes, there have been a lot of discussions between administrative record owners and Statistics Sweden (because of implementation of the new Data inspectorate act from EU). Taxauthority which is the base of the businesspopulation in sweden have had a discussion with statistics Sweden.
- Application of confidentiality rules should be done in a way to promote access to data while maintaining confidentiality.
- Contract between recordholder and authority is crucial even if the record is needed in statistical purposes.
- Official statistics in Sweden is always published aggregated so there is never any problem with disclosure. All published energy statistics uses statistical disclosure controls.

## **Statistical Quality**

## **Elements, Relevance, accuracy, timeliness and coherence**

### **Relevance**

Do the data elements that are being captured in the administrative record correlate to the concepts and definitions of the data user? Although it is often less expensive to mine administrative data than to collect the information via a survey, the analytical goals must be met with the administrative data for it to be a useful endeavor. Indicate the source, vintage, and how good definitions and classifications match to the survey data, and to the needs of data users.

### **Timeliness,**

Administrative data of this type is often unavailable until well after the reference period. In the case of using administrative data for a frame, it may be well out-of-date by the time it can be used. Additionally, if administrative data are integrated with survey data, it is important that the administrative data be as timely as the survey data: otherwise the entire process can be held up.

### **Accuracy**

Administrative data doesn't have the same quality check in edits as survey data does. Some edits are usually performed by the administrative organization, but their nature and purposes are usually different from those of the statistical agency. The result can be data quality issue when the statistical agency tries to use the administrative sources for statistical purposes, particularly with no or limited ability to recontact the originator of the information. Additionally, sampled administrative data may not adhere to any standard sampling scheme, introducing possible biases and making the calculation of sampling error difficult. Finally, if the administrative data are used as a frame in addition to or in place of another one obtained from data collection, it may not be possible to analyze the issues of coverage and nonresponse. On the positive side, many administrative data sources are censuses, meaning that there will be no sampling error in the estimates obtained from them. Indicate the contribution to key estimates from administrative data.

### **Coherence**

This is another significant consideration with administrative data. This type of data is typically captured for another purpose and, as a result, will not necessarily mesh/merge with already-defined concepts that might exist on other statistical holdings. This can be true in the case of concepts and definitions, and even in the sense of coverage and sample design. Administrative data might cover only a portion of the target population, making it problematic to use, or a sampling strategy may have been employed making the calculation of survey weights difficult to perform. There are cases where survey designers should have input into the design of the administrative systems, which can greatly increase the coherence of the data. List any exclusion that may complicate comparisons with other data. Indicators may include a measure of the target population not covered.