Techniques for Maintaining Frame Quality

May, 2017

When developing a sampling plan, the frame serves as the foundation on which the entire survey is built. Without an accurate and comprehensive survey frame, the survey estimates will not be as reliable and efficient as originally planned. For example, frame under-coverage can result from units missing from the frame, or miss classification (wrong ISIC code, falsely coded as inactive, etc.) that cause them to be excluded from the survey population. Over-coverage could result from duplicates on the frame, miss-classified units (again wrong ISIC code, falsely coded as active, etc.).

While some aspects of over-coverage can be detected during data collection, the under-coverage issues are more difficult to detect without an effort being made.

For these reasons, it is important to review the frame prior to designing and selecting the sample. In the case of a Business Register (BR), it is important to have regular reviews of its contents in order to ensure that it accurately reflects the business population required for the surveys.

Reviewing the Top Contributors

Reviewing the top contributors on the frame is an easy technique to ensure that that the largest units are present. For this, the frame is sorted by key measurement values (such as revenue and number of employees) and the largest units are reviewed by a Subject Matter expert. The goal is to ensure that the largest units are on the list, and that no errors have occurred to cause a small unit to appear on the list. This type of work requires a critical eye and knowledge of the business population.

This work should be done at the overall population level as well as other key levels such as industry, and perhaps some geographic breakdowns; in particular those that are used when releasing estimates. It can also be done at the different levels of business hierarchy such as Enterprise and Establishment.

It is likely appropriate to have different people review Top Contributors listings. For example, there may be Subject Matter experts for some industries that should review listings related to their area of expertise, while there may be personnel in satellite offices that should review listings by geographic area. It is important to match as many of the listings as possible to personnel who have good knowledge of the area.

This exercise not only helps to ensure that all large units are present on the BR, but also helps to ensure that units are properly classified based on key variables of importance. For example, if there is a large bottling company, the people reviewing the industry level reports would expect it to appear on a certain report with the geography review having the same expectations.

Reviewing the Top Contributors listings is a quick and easy way to ensure proper coverage of the largest units, which tend to account for significant portions of the estimates. Conducting the reviews at subnational levels ensures that smaller units that are significant contributors to a particular group of records are also well covered.

Reviewing Totals and Counts

Another quick method for assessing proper coverage is to review totals for key variables and counts. For this technique, key totals and counts are listed for review. These listings can be at the overall level, or at similar breakdowns as for the Top Contributors listings. Again, depending upon the listing, there may be different personnel better suited to review certain listings and judge whether they compare well to what is expected.

Some examples of totals and counts to include are:

- Overall total revenue and total number of employees,
- Total number of units within each industrial sector,
- Total number of units within geographical breakdowns,
- Total number of units for the various industry sectors by geographic level.

It is important to refer to the contents of the BR to determine the various counts and total breakdowns to review.

The knowledge for judging the accuracy of the numbers produced from the survey frame can come from experience in the area being reviewed, totals coming from other sources such as administrative files, or any other resources that might be available.

In general, if the frame is not reflecting an accurate picture of the population, neither will the estimates of any survey using it as a frame.

Reviewing Record Level Data

In addition to ensuring that totals and counts fall within expected limits, reviewing the micro-level data can also be done in an automated fashion. This type of work will take more time and expertize to set up, but once the initial investment is made, it will lead to a better quality frame with minimal work. The general idea of this check is to ensure that the values on the frame fall within certain ranges and that the relationships between variables are being maintained. For example, certain data items are restricted to a codeset; checks can be implemented to ensure that these items do not contain invalid values. Other checks can ensure that dates are falling within reasonable limits. A more complex example would include testing relationships such as holding companies are expected to have large assets, low revenue and no employees. These rules can be difficult to determine all at once, so it is expected that they be added through time. The key is to ensure that every rule gets documented in a central location, so that everyone working on the project knows which rules exist and what they are meant to detect.

These checks can be comparisons of values, percentiles or even graphs. Similar to the previous checks they can be conducted on the entire BR or subsets of the BR. Some examples of these checks are:

- Checks of codesets and dates:
    - Basic codeset values should be checked to ensure that they only contain valid values.
    - All dates should be earlier than the date of the checks being run. Is there a lower cut-off date for some variables? For example, if starting to include date stamps on the BR, there should not be any earlier than the date that the stamping was initiated.
- Check of thresholds:

    - Do some ISIC codes expect a minimum threshold of employees? For example, if an overnight fishing excursion (ISIC Rev4 = 5011) became miscoded to the accommodations sector (ISIC Rev 4=5510), such errors could be detected by reviewing the business names of the units in ISIC Rev 4=5510 with the fewest employees.
    - Do some ISIC codes expect a maximum number of employees? Should doctors' offices with more than 3 employees be reviewed for erroneous ISIC codes?
- Checks of relationships:
    - Graphing related values such as revenue against the number of employees, and looking for records that do not follow the general trend. The graphs can be produced by different industry sectors, or by particular ISIC codes, and can detect errors such as data capture and /or collection or ISIC coding.
    - When reviewing the entire BR, should certain industries account for the largest units (i.e., units below the 80th percentile for revenue should be reviewed for mining). Should other industries be smaller, such as farms expect to fall below the 20th percentile for the number of employees?
    - If parent and child relationships are reflected, the ParentID and flags indicting the level of the record can be compared for consistency.
- Large changes
    - If a quantitative value changes significantly since the previous snapshot, it should be checked. However, studies will need to be done to define a significant change.

It should be noted that when developing a BR, most of these types of tests of relationships will likely become part of the business rules that will contribute to the regular BR maintenance. This highlights one of the largest benefits of developing a centralized business register; most maintenance activities are done only once rather than individually for each survey.

Profiling and Updating the Largest Units

Giving the largest and most important units within the BR extra attention could go a long way towards building a relationship that will lead to extra co-operation from the businesses. One approach that could help with maintaining the BR would be to assign head office staff a certain group of businesses for which they will be the key contact person. The businesses could then be sent a letter advising them of the new procedures and inviting them to make use of their contact person should they have any questions about completing questionnaires, data requests or any other item for which they feel the statistical office might be of help. The letter would also advise the businesses that their liaison will also periodically contact them to ensure that their business is accurately reflected on the BR in order to facilitate efficient practices that are targeted at minimizing survey contacts.

With some creativity, a relationship that works well for both the NSO and the large businesses can be established. For example, perhaps during collection procedures of a key survey, the liaison officer could make an appointment to visit the businesses to help complete the survey questionnaire(s), negotiate the best ways to approach future survey collection, update the BR information, provide a tour of the NSOs website to show them where they may find information that would be useful for their businesses, etc. Once a relationship is built, procedures could be further expanded to collect identifiers for key administrative data sources in order to facilitate replacing survey data with administrative data. Based on what is available from the administrative sources, shorter survey questionnaires could be developed to collect only the remaining unknown data.

At StatCan there are two teams that are dedicated to these types of activities. One team has extra training in order to accurately represent the most complex and important structures on the BR and provide extra services to these most important businesses. The second team also specializes in this area, but act as the liaison for many more businesses and have less contact with each business. It should be noted that this type of structure is not necessary and the tasks of the liaison officers could be divided amongst all staff members in order to share the responsibilities, with the more experienced liaison officers taking on the more complex cases.

Building a relationship with businesses may require an initial investment of time and resources, but it should pay off through lowered collection costs in future years and more accurate reporting.

Routing Rules for Updates

To help ensure the quality of the data being applied to the BR, certain updates can be completed by only those with experience in a particular area. For example, if interviewers are permitted to apply updates to the BR, it may be beneficial to block them from updating certain fields, and request that they send such updates to their supervisors or highlight the updates on the questionnaires for review by the head office. When using output from a computer assisted interviewing tools (CAPI / CATI[1]) to transfer updates directly to the BR, the rules could be set up to have certain updates sent to listings for review before being applied to the BR.

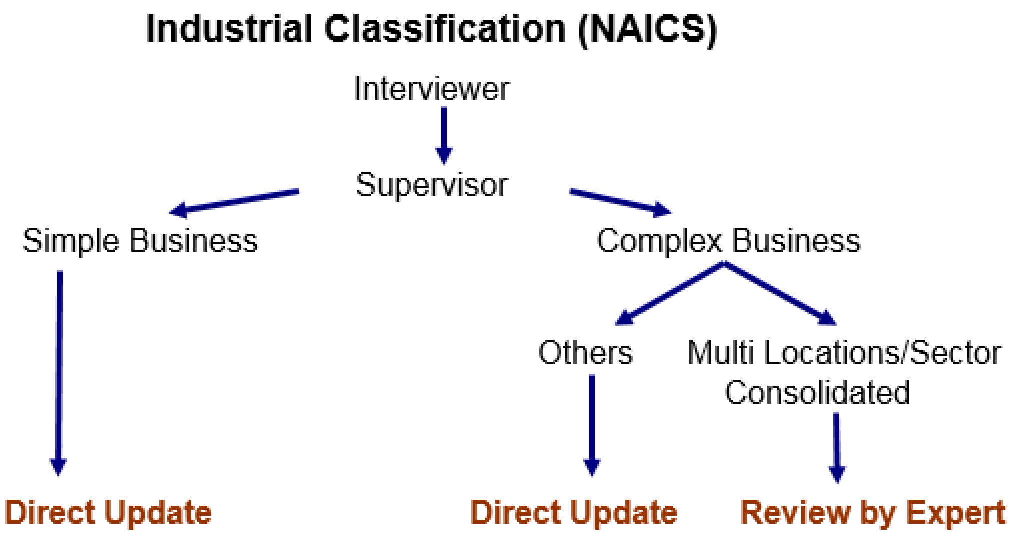
For example, at StatCan BR updates come from a variety of sources and depending upon the complexity and size of the business as well as the variable to be changed and the source, not all updates are automatically processed on the BR. Instead, certain changes need to be researched further and approved by a supervisor while others are sent to subject matter experts for review before being applied to the BR.

Below are two examples of how key updates are delegated within StatCan in order to achieve a balance between BR accuracy and timeliness, while limiting staff workload.

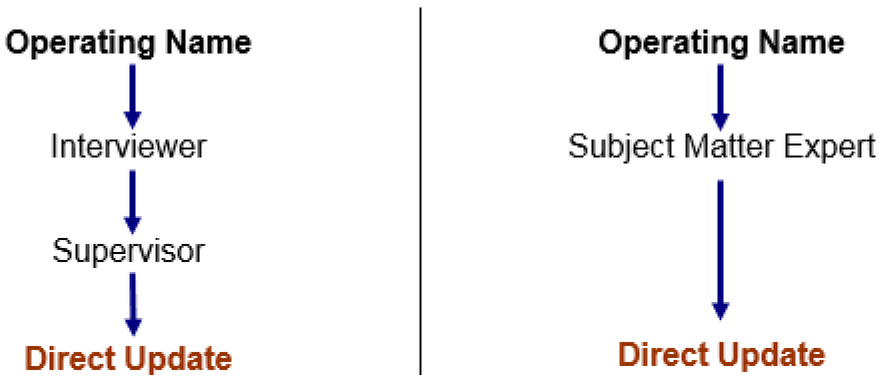
Example 1: Industrial Classification (NAICS) Update

When an interviewer changes an industrial classification, it automatically gets sent to a supervisor for verification. Once accepted by the supervisor, it may be automatically updated on the BR or sent to an expert for review, depending on the complexity of the business.

[1] CAPI refers to Computer Assisted Personal Interviews, while CATO refers to Computer Assisted Telephone Interviews.



Example 2: Operating Name Update
For this update, the complexity of the business is not considered, however, depending on the source of the message the update can be applied directly to the BR or sent for review. If the source is an interviewer, the update goes to the supervisor to review. If the source is a subject matter expert the update is applied automatically to the BR.



<u>Closing Remarks</u>
As mentioned at the beginning of this document, frames form the foundation on which surveys are built. Without a strong foundation, additional resources will need to be invested at most other stages of the survey to overcome these initial issues. Depending upon the frame issues that are encountered, the survey estimates could be inaccurate, biased or even not applicable thus rendering the whole exercise a waste of time and money. Because of this, it is important to make the proper investment to ensure that the survey frame is as complete and accurate as possible.