|  |  |
| --- | --- |
| **Business Survey Infrastructure Component** | **Summary of Discussions, Decisions and Future Activities**  **Business Register and Business Establishment Survey Mission # 3**  **April 19-22, 2016**  **Belmopan, Belize**  Prepared by Statistics Canada for the  Statistical Institute of Belize |

The third mission of the Business Survey Infrastructure component of the Project for the Regional Advancement of Statistics in the Caribbean (PRASC) took place in Belmopan, Belize from April 19-22, 2016 between Statistics Canada (StatCan) and the Statistical Institute of Belize (SIB). There were 10 representatives from SIB and 3 participants from StatCan.

The objective of this third mission was the development and specification of the edit and imputation procedures that will be used during post-collection to clean and complete the data received through the Part A questionnaire and the Part B series of questionnaires.

The agenda is attached in Appendix 1, the list of participants in Appendix 2, and the IT development requirements in Appendix 3.

1. The meeting was started on Tuesday with opening remarks from StatCan. The importance of the edit & imputation process from a quality perspective was emphasized, and it was stated that survey outputs are dependent on the respondents' interpretation of the concepts of interest. An up-to-date schedule for the week was reviewed with a focus on tangible deliverables: edit and imputation specifications. The agenda for the week and the related deliverables was accepted by all participants.

After this introduction, StatCan presented an overview of data processing. The presentation was followed by a context-specific review of the Business Establishment Survey (BES) requirements and SIB capacity for implementing automated solutions. The StatCan and SIB teams reviewed the projected workload for capturing, editing and imputing BES data based on sample size and expected response rate. Preliminary numbers of collection progress were discussed and the StatCan team proposed a formula for calculating the collection response rate. The teams also brainstormed a list of software and programming languages that could potentially be part of the SIB data processing tool box for the BES. These options will be evaluated to confirm the best fit in terms of cost/benefits for each process. Lastly, the SIB presented CSPro, the application that will be used to capture data reported on the Part B paper questionnaires. StatCan commented on the implemented edits and recommended changes.

1. On Wednesday, the StatCan team started by reviewing working assumptions with regards to SIB IT capacity for developing various components and user interfaces for processing BES data. A draft process flow model was proposed for discussion. The management of survey inputs was discussed as well as the data integration procedures for Business Register maintenance and for populating the survey database. The assumptions behind the draft process flow for data editing and imputation tasks were validated. The SIB IT team confirmed capacity for automating procedures for editing and deductive/rule based imputation.

A presentation on data editing was given by StatCan during which the advantages and disadvantages of manual editing as opposed to automated editing were discussed. The last part of the talk focussed on the decision tables as a tool to document and develop the edit rules needed for the BES. A concrete example of possible edit rules for the Business Revenue section of the Generic questionnaire was presented which helped SIB better understand how to properly use the decision logic tables. The presentation was followed by a group exercise where the edit rules were developed for the Wholesale/Retail questionnaire.

In the afternoon, StatCan presented examples of deductive/rule based imputation specifications and outlined the benefits of automating these procedures to ensure data are treated systematically, effectively and with consistency. The criteria for determining the usability of records were discussed and the specifications were drafted. The benefits of reviewing the result codes (refusal, unable to contact) and key variables, particularly for important contributors were discussed. Alternate process flows were defined for records to be accounted via reweighting (unusable active records) or imputed to zero (unusable inactive). Only usable records with reported total revenue will be edited and imputed through the principal process flow.

1. On Thursday, StatCan reviewed the editing specifications that SIB adjusted from the day before. These were related to the wholesale and retail questionnaire and to the validation of the response code. StatCan then presented the goals of imputation and the various methods of imputation, with their strengths and weaknesses. A brainstorming session followed the presentation in order to define the appropriate BES imputation strategy. Specifications for the chosen imputation strategy were documented. The level of automation was also discussed considering the development cost vs the benefits. During these activities, the process flow model was kept up-to-date considering the manual and automated imputation tasks, and the related imputation information structure. A second brainstorming session took place, this one under SIB's leadership, to identify the best IT solutions to support editing and imputation processes. A discussion on the number of out-of-scope units observed during collection resulted in an improved trust for the Business Register. Another discussion on the questionnaire completion rates now motivates a weekly discussion on the progress of the collection operations during project teleconferences.
2. On Friday, StatCan presented the proposed process flow model that was defined with the SIB team to review the working assumption and to assess the overall IT effort that will be required to automate the execution of functions. With support from StatCan, the SIB led a discussion to define the decision table for the transportation and communications questionnaire. StatCan defined the requirements for the Post E&I process and provided examples of procedures. The SIB team proposed system solutions for automating these functions. In the afternoon, the StatCan team presented the proposed process flow model to the SIB management team and summarized the benefits of automation. The IT development effort and execution timelines were estimated. StatCan summarized the joint accomplishments, and presented the future activities and recommendations for implementing solutions for data processing. Closing statements were presented by SIB and StatCan.

**Mission outcome:**

1. A BES process flow model was developed.
2. Training on edit and imputation methods was delivered, with focus on strengths, weaknesses and implementation complexity of each methods.
3. An edit and imputation strategy was developed.
4. Training on decision table was delivered in the context of statistical surveys.
5. Specifications using decision tables were developed for the Business Revenue module, the entire Wholesale & Retail questionnaire, and the entire Communication & Transportation questionnaire.
6. Specifications for the nearest neighbor imputation methodology were developed.

**Future activities:**

1. SIB will revise the hard edits currently implemented in the CSPro capture application by May 16, 2016 as follows:

* The systematic derivation of the total from the parts (in the accounting edits) will be removed.
* Blanks will be accepted rather than being initialized to zero for unreported cells.

While revising the application, SIB may want to extend the use of the "Other" category to capture the write-ins provided by respondents. SIB should also consider using another color to highlights failed accounting edits.

1. Up-to-date business information obtained from Part-A will be loaded to the Business Register by August 31, 2016.
2. SIB will finalize the record layout with a complete set of CSPro variable names as well as derived variables required for edit and imputation by May 29, 2016.
3. SIB will finalize decision tables for all 10 questionnaires by May 27, 2016. To reach this deadline, it is suggested that SIB statisticians complete two questionnaires every week and share with StatCan.
4. StatCan will review the decision tables as soon as they are received, or by June 10, 2016 at the latest.

1. SIB will document post-processing specifications by July 1st, 2016.
2. StatCan will review the post-processing specifications by July 8, 2016.

1. SIB will adjust the survey schedule, taking into account the workload, by May 27, 2016.
2. The most recent collection response rates will be sent to StatCan once a week to be discussed during teleconference. The current format will be extended with information on initiated contacts.
3. SIB will continue to effectively adapt the collection strategy to the actual situation in the field by considering alternate reporting solutions and workload assignments.
4. Should SIB require to prioritize follow-up activities, a random prioritization strategy should be used to ensure no bias is introduced.

1. StatCan will consider releasing a certificate for the active participation of SIB in the PRASC training by November 25, 2016. StatCan will also consider issuing person specific certifications per attended training.

**Topics for the next mission**:

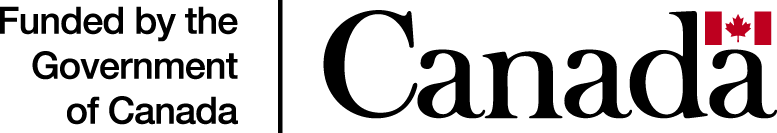
1. Discuss the development of a processing system to produce SUT estimates from survey data and the development of an application for balancing the SUT tables.
2. Discuss data processing techniques and systems for estimation, quality measures and disclosure control.
3. Discuss content of the preliminary report.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Leopold Perriott, Claude Poirier

Statistical Institute of Belize Statistics Canada

April 22, 2016 April 22, 2016



**Appendix 1 – Agenda**

|  |  |
| --- | --- |
|  | **Post Collection Data Processing**  Business Survey Infrastructure Component  **Belmopan, Belize**  **April 19-22, 2016** |

**Agenda**

**Objective:** To assist the Statistical Institute of Belize to set-up their post data collection processing systems. This includes a review of data processing requirements and the development of edit and imputation procedures.

**Participants:** Statistical Institute of Belize, Statistics Canada

**Tuesday, April 19**

**Afternoon (13:00-16:30)**

**Outline of the objectives/deliverables for the mission** (13h30-14h00)

**Overview of Data Processing Requirements –** Presentation and work session (14h00-17h00)

**Work session - Implementation plan**

* Review of response rate by stratum
* Demo of CSPro

**Wednesday, April 20**

**Morning (9:00-12:00)**

**Process Flow Model and Data Management** – Presentation and work session **(9:00-10:15)**

**Work session - Implementation plan**

* Review system solutions by process
  + Automated solutions
  + Manual solutions
* Define high level process flow model
* Draft preliminary ‘to do’ list

**Data Editing** – Presentation and work session **(10h30-12:00)**

**Work session - Implementation plan**

* Define edits for a questionnaire type using logical decision table
* Discuss manual and automated editing procedures for the BES

**LUNCH (12:00-13:00)**

**Afternoon (13:00-16:00)**

**Deductive and Rule Based Imputation** – Presentation and work session

**Work session - Implementation plan**

* Define deductive and rule-based imputation strategy for a part B questionnaire type by defining action on edits using logical decision table
* Define manual and automated intra-record imputation procedures for the BES

**Thursday, April 21**

**Morning (9:00-12:00)**

**Model and Donor Based Imputation** – Presentation and work session

**Work session – Implementation plan**

* Define a model or donor based imputation strategy for a part B questionnaire(s)

**LUNCH (12:00-13:00)**

**Afternoon (13:00-16:00)**

**Work session - Implementation plan**

* Define manual and automated model and/or donor based imputation procedures for the BES

**Friday, April 21**

**Morning (9:00-12:00)**

**Work sessions/Working groups:**

* Support for defining edits and deductive/rule-based imputation strategies for other part B questionnaires using logical decision tables
* Support for development of imputation procedures and specifications
* Support for development of system specification (i.e. prorate, donor selection)

**LUNCH (12:00-13:00)**

**Afternoon (13:00-16:00)**

* Review and finalize process flow model with subject matter and IT team
* Review recommendations
* Review summary of discussions
* Workshop evaluation
* Conclusion

**Appendix 2 – List of participants**

|  |  |  |  |
| --- | --- | --- | --- |
| **Invited participants** | | | |
| **Country** | **Organization** | **Participant name** | **Title** |
| Belize | Statistical Institute of Belize | Dr. Leopold Perriott | Director General, Office of the Director General |
| Mrs. Diana Castillo - Trejo | Deputy Director General  Office of the Director General |
| Mrs. Marilyn Pinelo - Lee | Manager, Economic Statistics Manager |
| Ms. Angelita Campbell | Statistician ll, Economic Statistics Department |  |
| Ms. Tiffany Vasquez | Statistician ll, Economic Statistics Department |  |
| Mr. Jefte Ochaeta | Statistician l, Economic Statistics Department |  |
| Ms. Jacqueline Sabal | Statistician ll, Economic Statistics Department |  |
| Mr. Gian Aguilar | Programmer Analyst, System Development and Data Processing |
| Mr. Melvyn Edenojie | Programmer Analyst, System Development and Data Processing |
| Mr. Darren Cal | University of Belize Intern,  Programmer, System Development and Data Processing |

|  |  |  |  |
| --- | --- | --- | --- |
| **Canadian participants** | | | |
| **Country** | **Organization** | **Participant name** | **Title** |
| Canada | Statistics Canada | Mr. Claude Poirier | Assistant Director, Business Survey Methods Division |
| Mr. Richard Laroche | Senior Methodologist, Business Survey Methods Division |
| Mr. François Couture | Data Processing and Analysis Officer, Enterprise Statistics Division |

**Appendix 3 - IT development and resources requirements**

Business Establishment Survey

* 10 questionnaire types
  + 5 generic modules
  + 17 industry specific modules
  + 319 distinct variables
  + 40-60 derived variables (estimated)
* Number of records
  + 4,100 sample
  + 2,000 usable records (estimated)
    - Require preliminary E&I (1200)
    - Require donor imputation (800)

Estimates of development effort by process

|  |  |  |
| --- | --- | --- |
| Process (E&I) | OPTION 1  Automated solution -  IT days | OPTION 2  Manual solution -Economic staff days |
| Preliminary E&I | 15 | 60 |
| Define imputation groups | 3 | n/a |
| Donor selection and imputation | 10 | 25 |
| Post E&I | 3 | n/a |
| Total | **31** | **85** |

High level BES deliverables calendar

* Delivery of final Post E&I dataset as input to SPSS – October 31, 2016
* Survey estimates – November 2016
* Balance SUT tables – 2017