

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION AND
DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Workshop on Statistical Metadata

(Geneva, Switzerland 5-7 October 2011)

Session 3: National implementations of the GSBPM

Topic II: Implementation of the GSBPM in development of Integrated Business Survey Program (Statistics Canada, Tim Dunstan and Alice Born)

I. Introduction

1. Along with other government departments, Statistics Canada has been asked to deliver its programs with fewer resources. Combined with the challenge of improving relevance while maintaining quality, this requirement will force Statistics Canada to achieve a much greater efficiency than at any time in the past. It is within this context that Statistics Canada is looking at significantly centralizing and standardizing the process involved in the production of statistical output. The proposal to develop an Integrated Business Statistics Program (IBSP) represents a way of achieving these objectives for business statistics.¹ The IBSP will be responsible for generating roughly one third of all Statistics Canada's surveys.

2. As part of this process, the IBSP decided to use the Generic Statistical Business Process Model (GSBPM) in its business analysis. This paper explains how the model was used and the lessons learned.

¹ Statistics Canada (2009): Integrated Business Statistics Program Blueprint, Enterprise Statistics Division, September 2009, Ottawa, Canada.

II. Statistics Canada and the Generic Statistical Business Process Model

3. The Joint UNECE / Eurostat / OECD Work Session on Statistical Metadata (METIS) developed the Generic Statistical Business Process Model (GSBPM) in 2009. The GSBPM version 3.1 was presented in Statistics Canada's Quality Guidelines (2009) as a tool to promote quality assurance in all phases of survey business process model.

4. The GSBPM (version 4) was formally approved by the Methods and Standards Committee as a reference model in March, 2010 to help guide programs in redesigning their business architectures and promote common terminology across surveys and statistical programs. The model is available on the Statistics Canada website in both official languages:

<http://www.statcan.gc.ca/concepts/gsbpm-msgpo-eng.htm>

<http://www.statcan.gc.ca/concepts/gsbpm-msgpo-fra.htm>

III. The GSBPM as a model for development of the Integrated Business Survey Program

5. In the spring of 2009, Enterprise Statistics Division (ESD) was tasked with developing a plan for a new approach to producing business statistics at Statistics Canada. This new approach was named the Integrated Business Statistics Program (IBSP). At its inception, it was agreed that the GSBPM would be followed in the development of a comprehensive business plan for the IBSP. The goals of the IBSP were to develop a model that allows, in the long term, the integration of as many business surveys as would make sense. However, for existing STC surveys, this sequential integration would happen over a period of 5 years and additional surveys would become candidates for integration as their systems age or, if for other reasons, (i.e., relevance, timeliness, potential for gaining operational efficiencies, etc.) they needed to be re-evaluated. Initially, this included 60 annual business (establishment-based) surveys covering the manufacturing, service and distributive trade sectors. Activity-based (i.e., agriculture and transportation) and enterprise-level surveys will be integrated later.

6. Higher flexibility entails complexity. It is therefore important to strike a balance between the two to ensure that a solid model, but a model that is relatively easy to understand, operate and maintain is developed. As such, it is important to establish at the onset that it would be impossible to develop a model to satisfy all of the particular requirements of each of the currently existing business surveys. Rather, the model must take the best approach for the majority and then, as programs become candidates for integration, an analysis must be conducted to determine how best the model can accommodate the new survey. It was agreed that the business model would be the GSBPM - unaltered at higher levels but customised at more detailed levels.

IV. IBSP Business Analysis

7. The goals of the IBSP business analysis are:

- develop metadata-driven systems
- optimize use of corporate services (e.g. collection, printing, methodology, etc.)

- maximize re-use
- minimize number of available tool kits (generalized systems and/or analytical tools)
- develop a data service center (central data storage)
- reduce micro-editing
- optimize use of electronic data reporting (EDR)
- standardize methods and processes
- use one single frame
- ensure data quality meets Agency standards
- balance between flexibility and simplicity

8. The IBSP business development team is using the GSBPM to help achieve many of these goals. Their approach was to map their existing business processes (i.e., IBSP Business Processes) to the GSBPM. Business use cases were then produced for each of the IBSP business process groups. Finally, a detailed business process model was produced by expanding the second level of the GSBPM and writing names and descriptions for the customized sub phases and sub-sub phases of the model.

IV. 1 Mapping the IBSP Business Processes to the GSBPM

9. The IBSP Business Process Development Team analysed which of the phases in GSBPM corresponded to their existing high level business processes. Eleven groupings representing the existing business processes were mapped to the 9 phases of the GSBPM (Figure 1). These IBSP Business Process Model groupings were used to divide Business Use Cases (BUCs) into manageable parts (for example, see Figure 2). While this exercise did not prove to be particularly helpful in retrospect as people continually confused IBSP groupings with Phases of the GSBPM, the approach is still maintained.

GSBPM		IBSP Business Process Model grouping																			
Expected # of workshops Level 1	Level 2 (process)	2	4	3	3	3	4	3	5	3	6	3	4	5	8	7	3	9	10	2	
		1	Design Content and Collection	Design Methodology	Build Collection Instrument	Design/Build/ Test System Component	Design/Build/ Test System Workflow	Sampling and Collection Processes	Post Collection Processing	Analyze	Disseminate	Archive									
1 Needs	1.1 Determine needs for information	x																			
	1.2 Consult and confirm need	x																			
	1.3 Establish output objectives	x																			
	1.4 Identify concepts	x																			
	1.5 Check data availability	x																			
	1.6 Prepare business case	x																			
2 Design	2.1 Design outputs																				
	2.2 Design variable descriptions		x																		
	2.3 Design data collection methodology		x																		
	2.4 Design frame & sample methodology			x																	
	2.5 Design statistical processing methodology			x																	
	2.6 Design production systems & workflow																				
3 Build	3.1 Build data collection instrument				x																
	3.2 Build or enhance process components					x															
	3.3 Configure workflows																				
	3.4 Test production system					x															
	3.5 Test statistical business process																				
	3.6 Finalize production system																				
4 Collect	4.1 Select sample																				
	4.2 Set up collection																				
	4.3 Run collection																				
	4.4 Finalize collection																				
5 Process	5.1 Integrate data																				
	5.2 Classify & code																				
	5.3 Review, validate & edit																				
	5.4 Input																				
	5.5 Derive new variables & statistical units																				
	5.6 Calculate weights																				
	5.7 Calculate aggregates																				
	5.8 Finalize data files																				
6 Analyze	6.1 Prepare draft output																				
	6.2 Validate outputs																				
	6.3 Scrutinize & explain																				
	6.4 Apply disclosure control																				
	6.5 Finalize outputs																				
7 Disseminate	7.1 Update output systems																				
	7.2 Produce dissemination products																				
	7.3 Manage release of dissemination products																				
	7.4 Promote dissemination products																				
8 Archive	7.5 Manage user support																				
	8.1 Define archive rules																				
	8.2 Manage archive repository																				
	8.3 Preserve data and associated metadata																				
9 Evaluate	8.4 Dispose of data & associated metadata																				
	9.1 Gather evaluation inputs																				
	9.2 Conduct evaluation																				
	9.3 Agree action plan																				
Start Date																					
Finish Date																					

Figure 1. UES Business Process Groupings mapped to the GSBPM

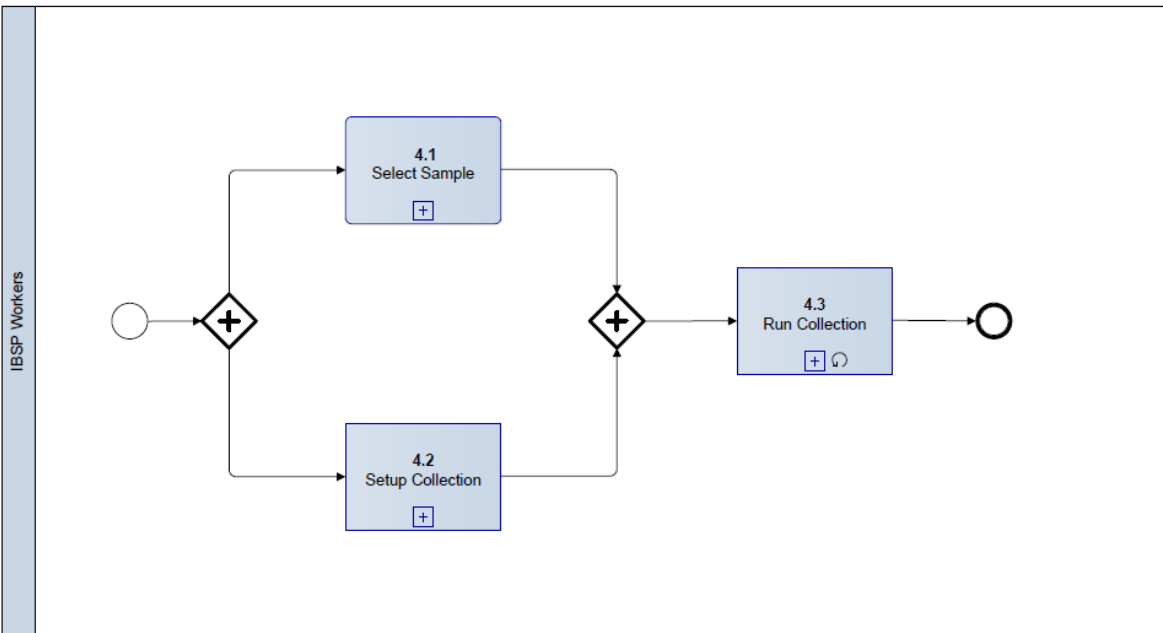


Figure 2. IBSP Business Process Grouping 7

IV. 2 Developing Business Use Cases

10. The business use case shows step-by-step how the system interacts with the actors, and what the system does in the use case. The business use case describes what happens inside the system, but not how or why. Business activities were determined for each BUC with regard to client or supplier/partner perspective. Figure 3 for illustrates BUC produced for IBSP Business Group 7 – Sample and Collection Processes, which corresponds to Phase 4 Collect of the GSBPM.

11. The information contained in the business use cases are used to develop the detailed business process analysis, and information analysis in the business architecture.

IV. 3 Detailed Business Process Analysis

12. The detailed business process analysis was carried out by expanding Level 2 of the GSBPM and creating another level. BUC descriptions of processes and activities were parsed and transformed into sub-sub-processes of the GSBPM (Figures 4, 5, and 6).

BUC#	BUC Name	Description
	<i>Name of the business use case</i>	<i>Brief description of what the BUC does from the perspective of the client and/or partner-supplier</i>
4.1	Select sample	This sub-process establishes the frame and selects the sample for this iteration of the collection, as specified in sub-process 2.4 (Design frame and sample methodology). It also includes the coordination of samples between instances of the same statistical business process (for example to manage overlap or rotation), and between different processes using a common frame or register (for example to manage overlap or to spread response burden). Quality assurance, approval and maintenance of the frame and the selected sample are also undertaken in this sub-process, though maintenance of underlying registers, from which frames for several statistical business processes are drawn, is treated as a separate business process. The sampling aspect of this sub-process is not usually relevant for processes based entirely on the use of pre-existing data sources (e.g. administrative data) as such processes generally create frames from the available data and then follow a census approach.
4.2	Set up collection	This sub-process ensures that the people, processes and technology are ready to collect data, in all modes as designed. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process. Where the process is repeated regularly, some (or all) of these activities may not be explicitly required for each iteration. For one-off and new processes, these activities can be lengthy. This sub-process includes: preparing a collection strategy training collection staff ensuring collection resources are available e.g. laptops configuring collection systems to request and receive the data; ensuring the security of data to be collected; preparing collection instruments (e.g. printing questionnaires, pre-filling them with existing data, loading questionnaires and data onto interviewers' computers etc.).
4.3	Run collection	This sub-process is where the collection is implemented, with the different collection instruments being used to collect the data. It includes the initial contact with providers and any subsequent follow-up or reminder actions. It records when and how providers were contacted, and whether they have responded. This sub-process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints. For administrative data, this process is brief: the provider is either contacted to send the data, or sends it as scheduled. When the collection meets its targets (usually based on response rates) the collection is closed and a report on the collection is produced.
4.4	Finalize collection	This sub-process includes loading the collected data and metadata into a suitable electronic environment for further processing in phase 5 (Process). It may include automatic data take-on, for example using optical character recognition tools to extract data from paper questionnaires, or converting the formats of data files received from other organizations. In cases where there is a physical data collection instrument, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material in conformance with the principles established in phase 8 (Archive).

Figure 3. Illustration of BUC for IBSP Group 7 Sampling and Collection Processes and GSBPM Phase 4 Collect

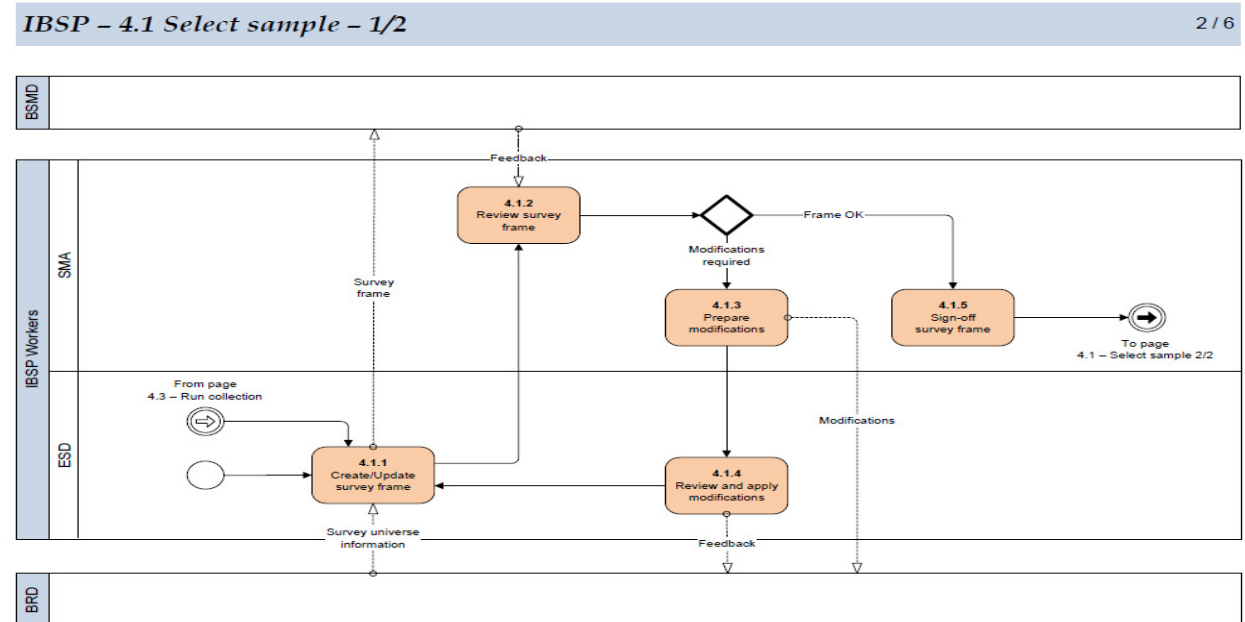


Figure 4 Example of detailed business process for IBSP Grouping 7 (GSBPM process 4.1 – Select sample)

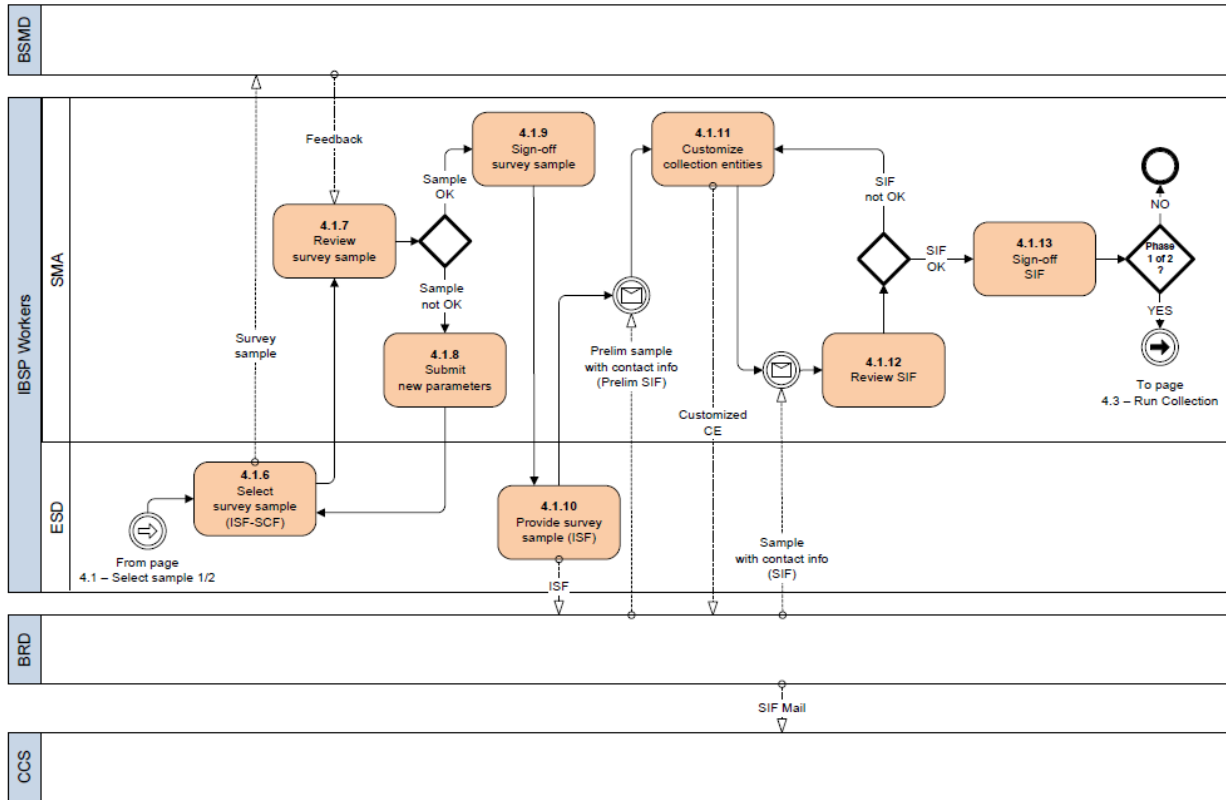


Figure 5. Example of detailed business process for ISBP Grouping 7 (GSBPM process 4.1 – Select sample)

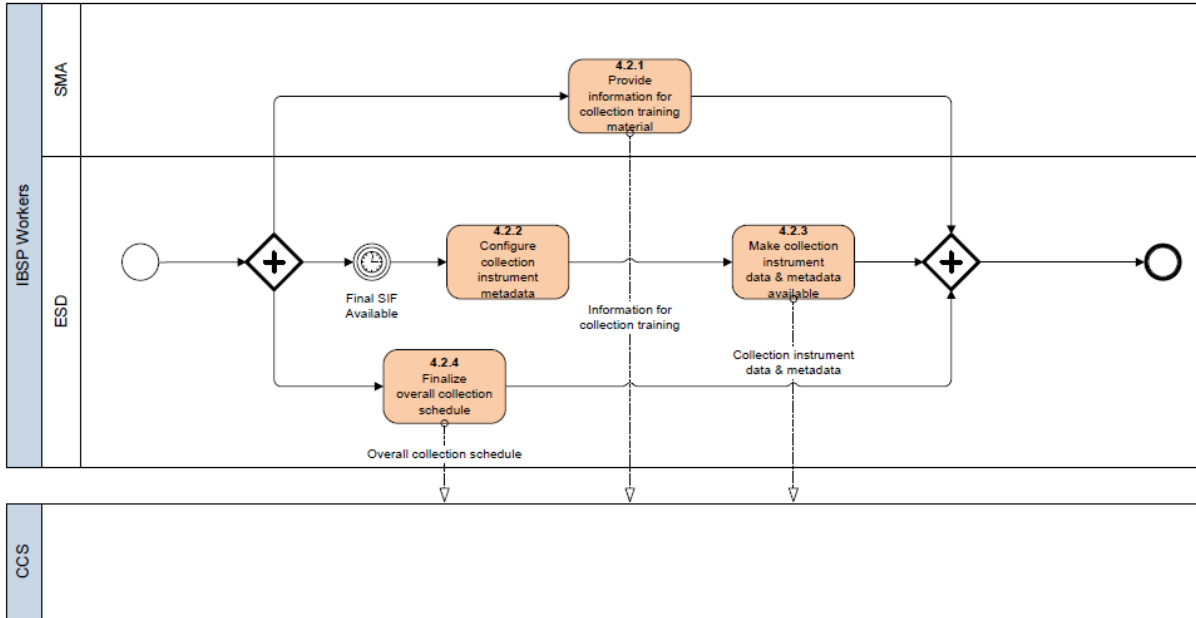


Figure 6 Example of detailed business process for ISBP Grouping 7 (GSBPM process 4.2 – Set up collection)

V. Information Analysis

13. Identifying all the information needs and flow throughout the processing is key to designing a robust, flexible, high-performance database. As a by-product of developing the BUCs and describing the activities for the Collection Phase of the GSBPM, the inputs and outputs of various sub-processes were listed. These then became information items linked to various parts of the collection process (see figures 7, 8, and 9 below).

BPM Task #	BPM Task Name	Input (I)
3.1 Build data collection instrument		
3.1.1	Provide survey collection instrument specifications	Specifications that would have been created in the design phase with the help of SM. - Final capture specs - Final edit specs The final output specs derived from these other specs above. Finalized questionnaire, including formatting, as metadata Locations from BR Negotiated content – between SMA and respondent
3.1.2	Prepare sample, testing data and expected outputs	Historical Q numbers/sample file Final output specs Final edit specs Final questionnaire Previous cycle test data and results Final capture specs
3.1.3	Update survey collection instrument specifications	Updated metadata
3.1.4	Evaluate output testing results	Test input file used by CCS Output file produced by CCS Output specifications that should have been used.

Figure 7. Inputs for GSBPM process 3.1 – Build data collection instrument

BPM Task #	BPM Task Name	Output (O)
3.1 Build data collection instrument		
3.1.1	Provide survey collection instrument specifications	<p>The final specifications (metadata) needed to build the E-Questionnaire (EQ), organized in a format that the Corporate Collection Service (CCS) can use.</p> <ul style="list-style-type: none"> - questions - flow pattern - capture specs - edit specs - other info required by CCS <p>The final specifications (Metadata) needed to build the questionnaire processing application (data capture system), organised in a format that he Corporate Collection Sertvice can use.</p> <ul style="list-style-type: none"> - capture specs - edit specs - output specs <p>Final specs to build large and complex Collection Instrument.</p> <ul style="list-style-type: none"> - Questionnaire Questions - Locations - Output specs for Cell capture <p>Finalized questionnaire including formatting as metadata.</p>
3.1.2	Prepare sample, testing data and expected outputs	<p>Sample or test mail file.</p> <p>Multiple test cases (sets of data) in electronic format designed to test every control feature.</p> <p>Expected outcomes.</p>
3.1.3	Update survey collection instrument specifications	New set of specs drawn from updated metadata.
3.1.4	Evaluate output testing results	Testing results evaluation (Reports that are the result of the test that can show differences that exist among the inputs)

Figure 8. Outputs for GSBPM process 3.1 – Build data collection instrument

Information Analysis

IBSP - 2.4 Design frame and sample methodology

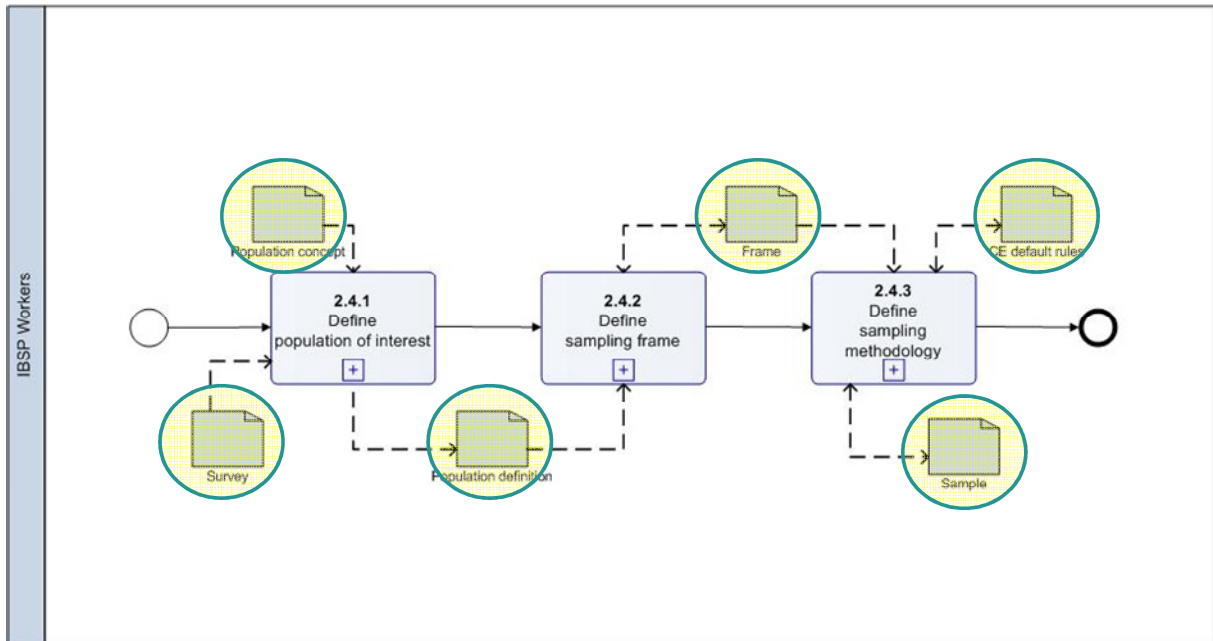


Figure 9. Determining information needs for GSBPM process 2.4 –Design frame and sample methodology

VI. Common Tools

14. Another by-product of doing a thorough detailed business analysis using a standard model such as the GSBPM is the need for tools can be determined and the number of tools can be minimised by making use of common tools. The IBSP has a set of common tools in use from the previous program. These tools can now be updated and revised to reflect the business needs generated by doing a thorough business analysis. Figure 10 below shows the tools and interfaces of the IBSP.

15. The common tools are:

- The Generalized Sampling software (GSAM) which provides basic survey sampling functions for survey developers. GSAM is a modular set of flexible programs that can be used to perform sampling functions such as stratification, sample size determination and allocation, sample selection and rotation, and sample and frame maintenance. The user has the option to use the modules separately or jointly, and interactively and iteratively as best suits the particular requirements.

- Banff which is a generalized system that offers methods of editing and imputing survey data in the form of SAS procedures. The procedures have been built in-house and behave exactly the same way as any other procedures of the SAS software
- The Generalized Estimation Software (GES) which is a SAS based system which aids in the production of estimates.
- Classification Management and Coding System (CMCS) is a generalized tool for accessing common reference files for coding data to standard classifications, and for interactive coding.

IBSP – Metadata Environment

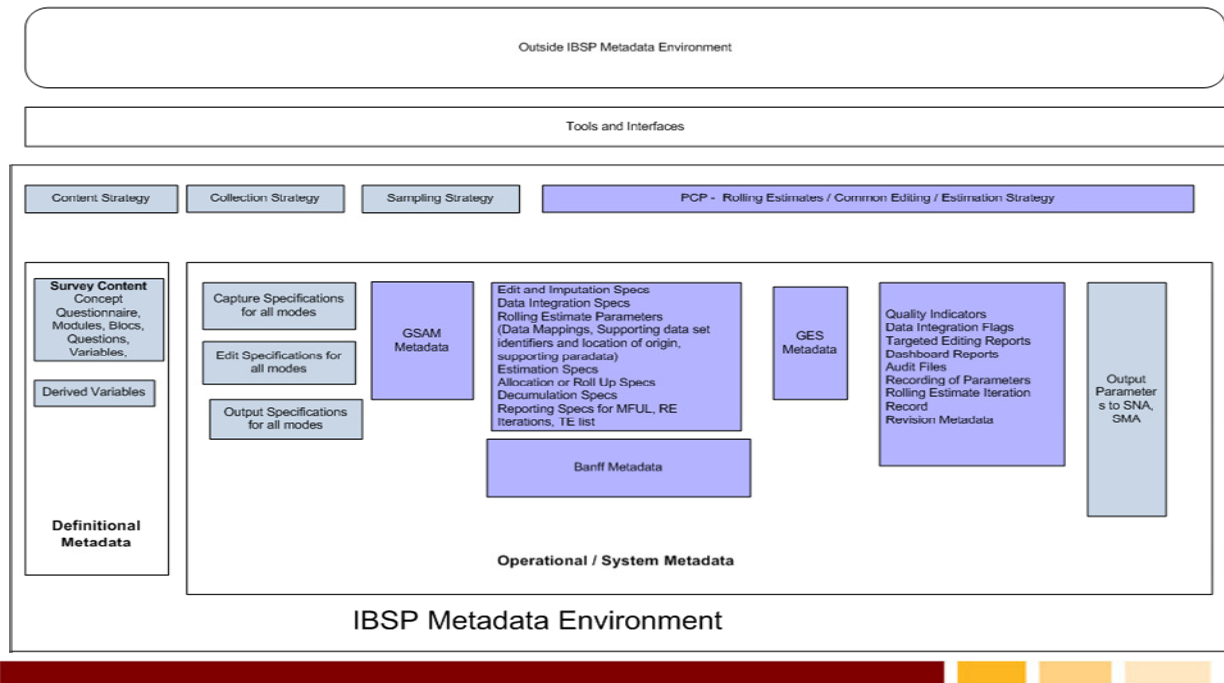


Figure 10. IBSP Tools and Interfaces

VII. Benefits of using the GSBPM in developing the IBSP business process

16. Benefits include

- The GSBPM is a well-structured high-level business process model facilitating description of coherent detailed standard business processes. Duplication and redundancy of processes is avoided by rigorous use of the GSBPM.
- The development of business activity descriptions was facilitated through use of the GSBPM.
- Description of information needs was the by-product of development of the business use cases. These needs can be applied to the entire business information model.
- Use of the GSBPM promoted the use of common tools and generalized systems fulfilling the business requirements of many of the processes.
- Documentation of the detailed business process was completed as the business process analysis advanced.
- Management of future changes in a process or sub process (Change Management) is greatly facilitated by use of the GSBPM.

VIII. Lessons Learned

17. Lessons learned include

- Mapping of the GSBPM to groupings of existing IBSP Business Processes was confusing and when either the GSBPM or the IBSP groupings were referred to people were often unclear about what was meant. In the future these mapping will not be done.
- The large group of people consulted at the beginning of the process was soon abandoned for a smaller more specialised group. It proved to be very difficult to reach consensus with the original large group. In the future smaller more specialised groups will be consulted for the conceptual work and the results will be brought a larger group for verification.
- Another lesson learned was to follow through with the momentum built at the beginning of developing the business process for a grouping, that is, once the business process model is complete, then as quickly as possible follow through with the task descriptions, and so on. Any delays in following through often lead to people forgetting where they left off and loss of continuum.

IX. Next steps

18. Information Architects are now developing information architecture for the IBSP including a Metadata Classification Model as well as a detailed Information Model. The Information Model includes semantic and logical models of information objects at level 0, 1, and 2 of the GSIM. In addition these objects are classified within the STC Metadata Classification Model. Generalised systems will be updated and developed based on the business requirements generated by these models. (see figure 11 metadata classification model)

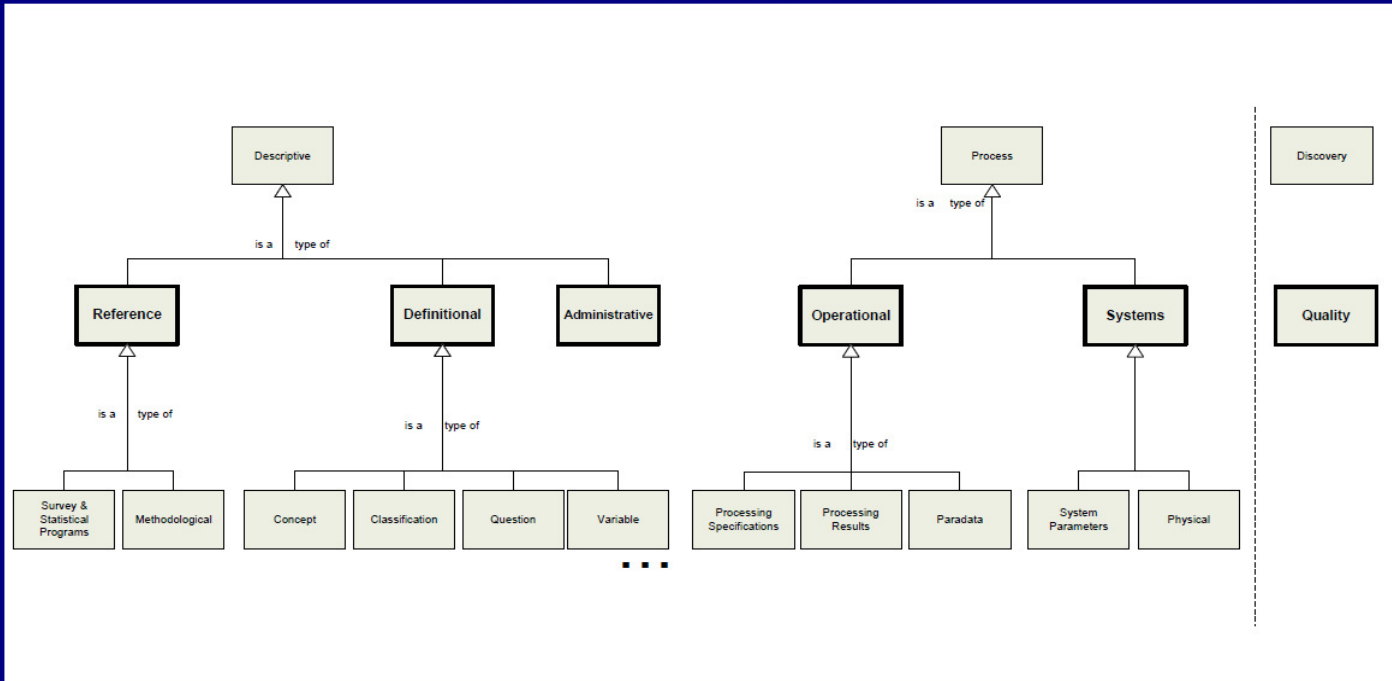


Figure 11 Draft Metadata Classification Model