CHAPTER 2. HOW TO CONDUCT A SURVEY

A. Introduction

2.1 This chapter provides some general information and rules on what needs to be done to conduct enterprise surveys. However, these procedures can also be applied to other business or household surveys if conducted for balance of payments purposes. Many countries continue to use a bank international transactions reporting system (ITRS); however, some countries are moving to survey-based collection systems. Individual countries decide on the general approach to adopt, but it is considered that a survey-based system is better able to collect the data for the balance of payments and IIP as transactions become increasingly complex while in some more "basic" circumstances the ITRS is still effectively used.

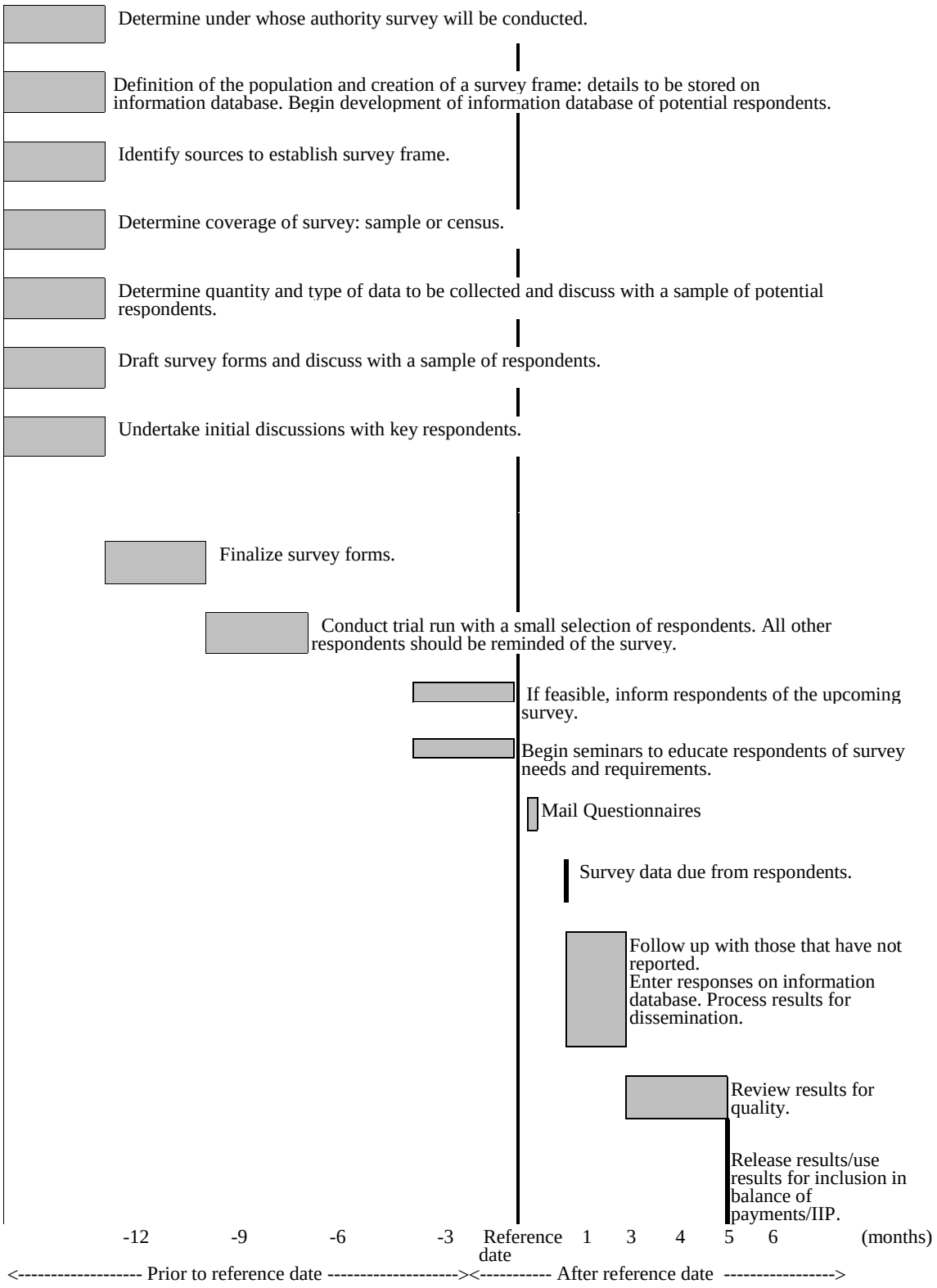
2.2 Enterprise surveys can be used to capture data for most elements of the balance of payments and IIP. They all have common operational aspects. This chapter begins with a description of the steps that are involved in conducting an enterprise survey. The general principles set out in this chapter are applicable in all cases, even though the collection process may range from data collection by telephone from a few large companies to highly organized, large scale, mail-based or web-based surveys. A survey may also be designed to capture a specific type of data or to obtain data that supplement other sources, such as an ITRS.

2.3 Surveys should be based on clearly defined objectives, sound collection methodology, and a well-established legal basis. Properly designed collection forms, full coverage of the population, well-defined data structures and classifications, and effective data validation and aggregation procedures are also required for enterprise surveys.

B. Timetable

2.4 Compiling agencies that have never conducted a survey will have a different timetable from those agencies that conduct them regularly. Box 2.1 below sets out a broad framework for a timetable, covering most of the aspects compilers will need to take into account when developing a survey. Each of the steps is elaborated further in this chapter. The timetable can be adapted to national circumstances but in general, the steps apply to most surveys. Agencies with more practice and experience, or for topics that are not as complicated as others, may not need as long a lead time prior to conducting the survey, and also may not need as much lead time for editing, processing, and disseminating responses to the survey.

Box 2.1. Draft Timetable for Conducting an Enterprise Survey



C. Responsibility for Producing Official Statistics

2.5 In many countries, a statistics act or formal legal arrangements exist under which the central bank or statistical agency has the authority to collect information or conduct a survey. In some jurisdictions, there may be no such legal authority. In those circumstances, proceeding with a collection on a voluntary basis may be appropriate, depending on the general relationship between the collecting agency and the enterprises being surveyed, but it is preferable to have legal authority as soon as possible. Good legal authority needs to state that reporting of statistical information is mandatory, especially for large enterprises.

2.6 In some economies, responsibility for collecting data for balance of payments purposes may be split between two or more agencies. For example, central banks may have responsibility for obtaining data from financial institutions, while the national statistical agency may have responsibility for the nonfinancial entities. In other economies, an investment approval agency or a financial supervisor may be a very important source of information about cross-border transactions. Whereas involving all the relevant agencies in the survey design will improve the relevance, as well as the overall coverage and accuracy, of the data eventually collected, it is important that the legal authority for the collection of these data allows all appropriate agencies to access the information. The survey questionnaire should make clear which agencies have access to the reported data. More information on institutional arrangements for compiling external sector statistics is presented in Chapter 8 of the *Guide*.

D. Creating or Updating a Survey Frame

2.7 The survey frame comprises the set of units subject to the survey and the details about those units that can facilitate conducting of the survey. The survey frame can be used to list units as well as to facilitate some of the steps involved in conducting the survey, notably through storing and tracking information on the units being surveyed.

Developing a register

2.8 The sources of information on potential survey respondents are varied, and the work required compiling a register to conduct a survey for the balance of payments purposes will depend on the extent to which a register already exists. A register is a set of records containing information on economic units that are included, or have the potential to be included, in balance of payments surveys.

2.9 If there is no existing register or if it is only rudimentary, these are some of the sources that could be used to build a register:

- Existing registers of enterprises maintained by the statistical agency or other government agencies for other purposes that might serve to provide useful information on those enterprises with international transactions or positions;

- Government administrative sources. Depending on legislation and administrative arrangements or the authority of the collection agency, these might include taxation records, files, or lists (including value added tax files and Customs files);
- Information held by foreign investment approval agencies or marketing boards;
- Information held by regulatory authorities (such as those responsible for supervision of financial institutions);
- Statutory company reports and company registration details;
- Records held in foreign exchange control or international transactions reporting systems;
- Media reports (e.g., business magazines, newspapers, or trade journals);
- Publicly available databases and reports such as the stock exchange register, commercial equity registry information services, international credit rating agencies' publications, market research reports or services by accounting or brokerage firms;
- Industry and trade associations. These associations can make available lists of members, often with indications of their financial size;
- Telephone directory.

2.10 All of these sources have limitations and should be used together. Moreover, the extent of their coverage may be larger, smaller, or different from those for statistical purposes. However, most of them will help build a register, such as by providing information on the size of enterprise activity. For example, a list of importers from international trade statistics may classify importers by size of importing activity in a specific time period.

2.11 The development of the balance of payments register may be regarded as a two-part activity. In the first phase, enterprises with potential balance of payments transactions are identified as being engaged in cross-border activities via sources previously described. The register then becomes the source list for enterprises to be included in balance of payments surveys. These enterprises are then compared with units already listed on the register. The compiler should make every effort to identify all units with potentially significant balance of payments transactions.

2.12 In the second phase, more information is obtained on enterprises that were identified from initial sources and are not yet in the register. It is unlikely that all enterprises identified will be entered in the balance of payments register because some units will not be engaged in activities of interest. An exploratory survey may be used to discover what, if any, balance of payments activities the enterprises are involved in and the size of those activities. Model

form 1 in Appendix 8 is an example of a form that could be used for an exploratory survey. In such surveys, the form should be kept simple so that collection and processing costs are minimized. Implementation of the exploratory survey may reveal problems with source data; for example, a number of duplicates may be found, or it may be difficult, in practice, to identify all units listed by a source.

2.13 Enterprises determined from the exploratory survey not to be engaged in balance of payments activity should be recorded on a list of residual enterprises and monitored. These enterprises should be approached in future exploratory surveys approximately once every five years.

2.14 A threshold can be established for including enterprises in the exploratory survey—particularly if source lists are large. (However, the compiler may wish to add units of particular significance directly to the register, rather than including them in the exploratory survey, so that these units participate in balance of payments collections as soon as possible.) Units below the threshold may be recorded on supplementary lists.

2.15 If units in the balance of payments register are stored at the enterprise group level, the register should include some information on company group structures. This information is important if an enterprise is identified as potentially suitable for inclusion in the balance of payments register. In these cases, the compiler must know whether or not the enterprise is part of a group that has already been identified. An enterprise register for an international transactions reporting system (ITRS) could also record information on a group basis even though the collection (survey) may use the enterprise as the statistical unit.

2.16 Once the balance of payments register is established, it should be updated and extended as necessary. The compiler should keep abreast of developments taking place in the economy—for example, by reading the financial news. He or she should also be alert to any major changes among sources used for initial identification of enterprises with potential balance of payments transactions. Some enterprises will need to be dropped from the register as they do not have the type of transactions/positions being measured, while others will need to be added.

2.17 The register should include unit name, address, contact officer, telephone and facsimile numbers, and area and size of balance of payments activity. While the balance of payments register is a logical database, it need not be physically separated from other statistical registers. For example, some national statistical offices maintain a single register for all surveys, including balance of payments collections. Model form 2 in Appendix 8 is an example of a form that could be used to collect data to build up a register.

2.18 Periodic references to the sources above are necessary to keep the register current. Some enterprises may be surveyed for a variety of topics, while others may be surveyed for only one. Some will be covered in an ongoing survey; some only in a benchmark survey or in census conducted infrequently. As a consequence, the register of companies will need to cover many of the elements of the system.

2.19 Register's sources should be reviewed progressively, and lists of units newly identified from these sources should be checked against the existing register. Unless more immediate action is required, information on newly identified units could be obtained by including them in the next exploratory survey.

Building the survey population

2.20 After the development of the register, the compiler needs to produce a list of potential respondents. A decision needs to be made as early in the exercise as possible as to whether to undertake a census, or compile data from as large a sample survey as possible. In determining the reporting population, various approaches are possible. In practice, compilers in many countries use a combination of two or three approaches when collecting data from enterprises:

- *Census:* Includes all members of the population;

- *Partial coverage collection survey:* Includes all enterprises above a certain threshold measured in terms of their dimensions (e.g. nominal capital) or other variables (e.g., significant cross-border activity);

- *Random sample survey:* Includes enterprises that are preferably selected according to rigorous sampling procedures, with the results "grossed" up for the whole population;

- *Stratified random sample:* groups population components according to the size of selected activity so that enterprises within different strata have different probabilities of selection. Usually, this is a combination of the partial coverage and random sample options but is more sophisticated and might produce a high level of coverage while remaining relatively cost-effective.

2.21 A census is useful to provide the benchmark for estimating the universe in subsequent surveys when samples may be used. However, for most activities, it is not necessary that the compiling country conducts a census covering all enterprises as it would be too much burden and many countries are not equipped for it. Maintaining an up-to-date information database of all additions (and deletions) of enterprises in the register is an essential part of keeping the estimates as accurate as possible.

2.22 For countries that have not undertaken an enterprise survey before, it will be highly beneficial to initially focus on the largest firms in their economies that are involved in the activity being measured, with less attention given to smaller firms. Consideration should be given to more sophisticated methods of compiling data, only after experience has been gained in conducting the survey, such as by conducting a sample survey with estimation for nonsample firms. Undertaking a sample survey without a good understanding of the relative size and importance of the enterprises being surveyed may produce data that cannot be reliably grossed up to a universe total (more details on grossing up techniques are presented below).

2.23 Grossing up the data without a census is more difficult than conducting a survey, which requires prior knowledge of the approximate size of the universe. The size of the universe involves two major dimensions: the number of entities in the universe, and the individual weight of each enterprise's transactions/positions. As economic statistics are primarily concerned with values, in any survey, the focus of a survey should be on those entities with the highest weights. In this regard, it may be appropriate to conduct a census of those enterprises that constitute, for example, 90 percent of the total activity/position being targeted and to conduct a sample survey or use models to estimate the remaining 10 percent of data. However, it is also important to stress that there are increasing demands for data on small- and medium-sized enterprises (SMEs), so that, if a sample survey is employed for those enterprises with the smallest contribution to the total on, it may be useful to bear this information in mind when the sample is designed, so that the detail on the SMEs is sufficiently robust for analysis, especially at the industry level.

2.24 The exploratory survey can be used to collect broad information on the size of balance of payments transactions of individual transactors in the population frame. This information could be used directly to estimate the impact of units not surveyed in the partial coverage collection, or it could be used indirectly in the creation of a framework for a sample survey. Because sample surveys are relatively inexpensive, they can be conducted at frequent intervals—for example, annually or quarterly—and can provide high quality factors for expanding results from partial coverage surveys. Another advantage with sample surveys is that the impact of non-response may be reduced. The disadvantage with sample surveys is the presence of sampling error.

2.25 Nonetheless, non-response is likely to be a concern with either a census or a sample survey. An appropriate approach to dealing with it should be decided before the data are collected (see paragraphs 2.56 – 2.62), and applied in a flexible way as the survey develops.

2.26 Large transactors must be approached each time a partial coverage collection is conducted. It is therefore important that the list of large transactors be kept up-to-date. Use of partial coverage collections can provide cost savings without much loss of quality. If the partial coverage approach is used for balance of payments surveys, the compiler should develop methods to measure, at frequent intervals, the contribution to balance of payments activity of all members of the population. These measurements could be made by using an exploratory survey, a sample survey of smaller units, or a benchmark census.

2.27 Efficient sampling procedures seek to keep both the number of units selected and the sampling error to a minimum. These objectives are usually achieved by stratifying the population. Two factors predominate in the determination of sampling error. One is sample size; the larger the sample, the smaller the sampling error. The other is the variability of the activity being measured; the wider the dispersion of the activity, the greater the sampling error. Population size is not an important factor unless the population is very small or the sample size approaches the size of the population. Stratification involves grouping the units into similar size bands and selecting an independent sample in each band. Variability of units in each band is less than the variability of the population of units as a whole. Typically all

units in the largest size stratum are enumerated. By using stratification techniques, the compiler essentially increases sample size for units likely to have large absolute variability in size of activity

2.28 In addition to supplementing partial coverage surveys, sample surveys can also be used as the primary survey method. For example, the compiler can use sample surveys as the principal source of information on international trade in selected services. However, compilers in many countries choose not to use a sample survey approach to supplement partial coverage surveys. Instead, they use benchmark censuses to establish the contribution of smaller units. These censuses are usually costly and undertaken infrequently. Therefore, revisions to results may be made at greater intervals. However, benchmark censuses typically provide more detailed information than other approaches and also establish whether or not some entities, which should now be included in the completely enumerated partial coverage survey, have changed size during inter-census periods.

2.29 The compiler must be careful to select the correct population. To do so, the appropriate target group, probably the larger enterprises, should be engaged through consultation meetings. These discussions should make these entities aware of the purpose of the survey and help the statistical agency design the survey so that it is most efficient in obtaining the desired information. Even where the statistical agency has the legal right to collect the data, this does not guarantee the cooperation of the target group – and cooperation is essential for good results.

2.30 In approaching the target group, the compiler needs to know not only the concepts that are to be measured but also the nature of the business activities that are being surveyed. He or she also needs to be aware of such things as the terminology used in the business activity, the nature of the operations, record keeping, and accounting practices of the target businesses in order to be able to communicate with the target group and to gain their respect and cooperation. Businesses are not all structured in the same manner. The information sought may be recorded in different ways in different organizations – especially for large complex entities – so some flexibility in how the data are captured is helpful.

Information content of the survey frame

2.31 Depending on the topic, the survey frame should be developed well in advance of the reference date (for more complicated topics, a lead time of twelve or even eighteen months may be appropriate).

2.32 The survey frame should include relevant information on each respondent and may serve as a tracking system for contacts with respondents. Relevant information to maintain on the respondent includes:

- Respondent name;
- Respondent ID code/registry or business number;
- Respondent address;
- Respondent contact person/authorized person to sign form;

- Designation of contact person/authorized person;
- Phone number of contact person/authorized person;
- Facsimile number and email address of contact person/authorized person;
- Activity code of respondent (i.e., business sector of respondent);
- Financial/fiscal year of respondent's enterprise (i.e., accounting period).

Survey log

2.33 A well integrated data collection provides comprehensive response logs containing information about the respondent and to track response status.
Response log details:

- Date letters/questionnaires are sent (could be automatically entered by computer);
- Date due;
- Response status (responded, liquidated, no response, respondent owned by another enterprise which will respond);
- Date first response received;
- Date of first and subsequent follow-up and nature of discussion and outcome;
- Edit date – case edited. Ensure follow up, as appropriate.

Computerizing the survey frame

2.34 Computerizing the survey frame may save compilers time and effort and reduce the scope for error in dealing with survey respondents. Careful maintenance of the database is also required, e.g., updating address changes, company name changes, etc. The type of issues and tasks that national compilers need to consider include:

- Taking account of the information flows to and from the database (the use of charts and diagrams may be useful in the planning phase);
- Taking account of the specifications required to query the frame and to generate reports;
- Designing a coding structure that embodies important defining characteristics of each respondent. This allows the sorting and analysis of respondents and the tracking of nonresponse;

- Ensuring that supporting hardware/software is sufficient for the task, for instance, that the memory and processing capabilities are such that response and retrieval times are acceptable;

- Allowing time for a thorough tryout of the system before "live" use;

- Backing up copies of data and software on a regular basis and storing them both on-site and at a remote location so that the files can be restored in the event of a system breakdown;

- Ensuring appropriate systems security and access authorization;

- Producing thorough documentation on the system; and

- Providing for a suitable filing system for associated papers, not least the completed survey forms.

Use of the survey frame

2.35 Once the survey is under way, the benefits of an efficient computerized survey frame become apparent. The benefits accrue both at the initial stage of distributing the survey forms and at the follow-up stage and beyond. Some of these benefits are presented below to provide the compiler with an idea of the capabilities of a computerized information database.

2.36 At the initial stage, compilers need to prepare and dispatch (either electronically or via the mail service) the survey forms and companion transmittal letters. The compiler can generate the mailing labels from the computerized information database. At the same time, a record can be made noting which survey respondents will be sent a questionnaire (e.g., mail status to “yes”). The compiler can also perform a quality check to ensure that the right respondent is receiving the right survey form(s).

2.37 This is particularly important if the mailing includes more than one type of survey form. If the statistical agency has never previously conducted such a survey, it may be useful to include with the survey form a postcard (or use some similar technique) that respondents use to acknowledge receipt of the questionnaire and identify the person to contact. In this way, compilers can quickly identify problems and initiate follow-up action. The national compiler should note, in the response log, those entities that have returned the postcard. The compiler should record the return of the completed survey form in the response log, along with the date of receipt, and any changes required to the information database (for instance, name, address, etc.). The compiler could run a report to verify that the correct entries have been made. Thereafter, he or she could distribute survey forms to the appropriate person for data capture and editing.

2.38 For a number of reasons, respondents may wish to contact the compiler using the contact information printed on the questionnaires. The compiler should keep a record of

significant phone calls and/or correspondence (such as requests for extension of the survey due date) either in a manual file or as a note against the appropriate record in the response log file, along with a notation on how the matter was resolved. If a survey respondent requires more time to complete the survey form, the compiler should note this information so that reminders are not sent. In such a case, the response log status is changed to indicate "in contact," and follow-up procedures are suspended for a specified period. If nothing is received from the respondent by the agreed date, the system should indicate this and prompt the compiler to re-contact the respondent. When new copies of the survey forms are requested, a re-mail is initiated, and additional forms are sent to the respondent with the follow-up procedures kept in place. The compiler must update the information database if updated information is provided and run a check to confirm that correct details have been stored.

2.39 At the planning stage, it is important to consider which fields will be automatically completed by the computerized system. For example, would the "date mailed" field be updated directly by the mailing program? Provide sufficient space for addresses, etc.; it may be appropriate to subdivide fields for some items (for example, distinguishing the postal code from the remainder of the address). Care needs to be taken when assigning default values to fields and (as a safeguard against inputting errors) have an accessible, on-line, separate file that lists those types of input that are currently acceptable for a specified field, so that unacceptable inputs will be rejected, for instance, entering an "alpha" value in a field where only numeric values are appropriate. It is also useful to have a comment area for the compiler's use.

2.40 When the closing date for the return of completed survey forms has passed, the compiler can identify the overdue survey respondents from the response log, based on the response status of "outstanding" and the mailing date, and prepare labels for envelopes and follow-up documents.

2.41 A number of reports can be produced on a regular basis to assess the status of the survey:

- *Transaction reports*: lists of changes to records sorted by name, ID code, date, etc.

- *Response log reports*: summary counts of survey forms mailed, received, and percentage outstanding; response log listing all survey respondents, survey status sorted by name, ID code, etc.

E. Draft Survey Questionnaires

2.42 The model survey forms provided in Appendix 8 could serve as a starting point for the questionnaires to be used to collect the data and could be modified for local circumstances.

2.43 The basis for much of the information to be collected-especially for financial data-should be available from the business accounts (such as its income and expenditure statement, the statement of changes in financial position, and the balance sheet), and the information collected should consider the accounting standards that respondents routinely follow in preparing and maintaining their records. However, as some of the information needed for the international accounts is not necessarily standard for drawing up the accounts, respondents may not have ready access to the data, or the data may require combining (or even disaggregating) data in the underlying management information system. Contact with respondents is, therefore, essential to ensure that there is sufficient lead time for them to put in place the systems needed to extract the data.

2.44 Balance of payments and IIP concepts and definitions need to be considered when designing surveys, and carefully compared to the concepts and definitions used in surveys. For instance, identifying residents and nonresidents may be difficult. The compiler may need to advise surveyors on suitable questionnaire design (for instance, to ensure that the relevant components of remittances can be estimated).

2.45 The questionnaire should indicate under what authority the information is being collected, that the information is for statistical purposes, and that no confidential information will be divulged without the consent of the individual respondent to which this information pertains and only aggregated data will be published. Commercial and personal data are available only to compilers of balance of payments and related statistics. The questionnaire and the transmittal letter should also indicate the reason that the information is being collected. As well, the questionnaire should provide advice on how the information is to be provided, setting out the measurement date, the currency of valuation, the valuation principles, as well as some additional information (such as the contact person's details, and the industry of the entity).

F. Initial Discussions with Key Respondents

2.46 The following is an approximate timetable for conducting a survey for the first time. Once a survey is being conducted on a regular basis, some of these steps may not be necessary or the lead time may not need to be so long.

About 12 months before the reference date of the survey:

2.47 Having identified many of the major enterprises to survey, the compiler needs to complete the development of an appropriate survey form, which should be field tested with the key respondents to (i) inform them that the survey will be undertaken in the coming year or so, and (ii) to give the respondents an opportunity to provide comments on the survey form and to ask questions.

2.48 Having face-to-face discussions with key respondents is a very efficient way of eliminating potential reporting problems, thereby limiting the extent of the need for follow up and raising the quality of the data. As was mentioned earlier, in surveys for the compilation of economic statistics, accounting records are central. It is often the case that accountants and economic statisticians use the same terminology, yet the meanings may not be the same.

Equally, economic statisticians may use terms (such as "financial intermediation services indirectly measured", "trade credits", or "foreign direct investment") that may not be understood in accounting terminology. Such matters can be clarified and the draft reporting instructions improved.

2.49 The survey form needs to identify the agency responsible for conducting the survey and under which authority the data are being collected, including sanctions for non-reporting, as relevant.

About 6 months before the reference date of the survey:
Trial run, if feasible

2.50 Conducting a small scale trial run with a sample of respondents no later than about six months before the reference date (even earlier may be useful) may provide many benefits, resources permitting. It may highlight where respondents have problems interpreting the questionnaire, and it may also serve to test the compiler's processing system. Highlighting and addressing problems at this stage will reduce problems at a later, and more crucial, stage.

About 3 months before the reference date of the survey:
Advance notification to identified respondents, if feasible

2.51 About 3 months before the reference date, resources permitting, the compiler could send a notification to all identified respondents to advise them that the survey will be conducted, the reference date, and a summary of the information that will be requested.

Training seminars, if feasible

2.52 In addition to having one-on-one or small group meetings with some key respondents, resources permitting, it may be useful for the compiler to arrange larger seminars for respondents. These seminars could review the survey form and identify any areas where respondents may not be familiar (such as the concept of *residence* or *center of predominant economic interest*). The point should be made that the information provided is confidential and only aggregated data will be published. The seminars also help public relations, and allow the compiler to advise respondents of the reasons for the survey. The seminars should serve as another opportunity to fine tune the questionnaire.

About 1 month before the reference date:
Re-contact respondents to remind them of the survey

2.53 It may help the response rate (at least, initially), if respondents are contacted (by mail or email) to remind them of the survey's reference date and when data are due to be reported.

During the first month after the reference date:
Mail out survey, along with acknowledgement cards, if feasible

2.54 As part of the questionnaire that is sent to the respondents, resources permitting, it may be helpful to include acknowledgement cards, which respondents should return to the compiler, to indicate that the survey had been received. Respondents should be asked to

return these cards immediately upon their receipt. For responses that have been received, this should be entered on the information database. For those requiring follow up, this should also be noted on the information database.

Between 1 and 3 months after the reference date of the survey:
Reporting date by respondents

2.55 It is recommended that data for an initial survey be requested within one to three months after the reference date, depending on the complexity of the survey and how readily respondents can access the information from the management information systems. If it is later than this, respondents may forget to complete and return the questionnaire. After the lapse of this time, for those respondents that have not reported, the compiler should follow-up more or less immediately after the due date. Responses that have been received should be recorded on the information database and made ready for processing. The information database should also note cases where follow-up is needed, either due to nonresponse or due to an inquiry or response that requires an action by the compiler.

G. Low Coverage or Low Response Rates

2.56 Hopefully a substantial share of the requested data will be received within the time allotted for the collection of the data, although there are likely to be some responses that remain outstanding, particularly if the questionnaires were not delivered and collected by enumerators. In order to obtain data that approximate the universe, there are various ways to estimate for low response or low coverage rates.

2.57 For countries that have conducted surveys previously and prepared universe estimates of positions, the previous estimates can be used as a starting point. For example, if a survey requested data from 100 enterprises, and by the cut-off date, returns have been obtained from only 70 enterprises, the compiler has to estimate the data for the missing 30 enterprises on the basis of the most recently reported data for these enterprises. These estimations can be calculated as followed: the changes reported by the remaining 70 enterprises that filed surveys are considered in relation to the volume percentage of the enterprises, reported vis-a-vis not reported, from the previous period. This technique can be refined by analyzing the changes by sector. If there has been a history of making estimates that have subsequently been revised, and if there is a consistent upward or downward bias in the initial estimates, the compiler should consider this bias into account when calculating the estimates.

2.58 If the survey is not a census, the results should be grossed up to produce a universe estimate, perhaps using an earlier benchmark survey or census as the basis for the grossing up factors. When the next census or benchmark survey is conducted, it may be found that the weights for the unsurveyed portion of the universe have changed. In that case, the data for the periods between the census/benchmark surveys should be revised to reflect this. The change in weights of the unsurveyed enterprises between the two censuses/benchmark surveys should be introduced gradually over the period for which the data are being revised, rather than introduce the change in weights all at once. Further details regarding grossing-up and estimating data appears in Chapter 8.

2.59 Other adjustments may include: taking account of exchange rate and price movements (especially for surveys of financial assets and liabilities); changes in financial markets that might affect interest and other income flows; for goods and services, adjusting the nominal values to volume measures (for example, through price deflators) to see whether the nominal values are consistent with the volume measures; noting developments in various types of markets (such as commodities and financial) to ensure that any changes are adequately picked up in the underlying data or to adjust the collection vehicle's questions accordingly; where there have been changes in legal or institutional arrangements (such as changes in exchange control or opening of segments of the local market to nonresident activity – either through direct sales or through permission to make investments) that may result in changes in the nature of cross-border transactions and positions.

2.60 Such approaches may provide reasonably good estimates for positions data. The methods used to derive universe estimates, when responses have been less than 100 percent, should be described in metadata.

2.61 However, where response rates are low for a survey that has not been previously conducted, alternative approaches may be required. For example, if the survey requested data from 100 enterprises, and by the cut-off date, returns have been obtained from only 30 enterprises, in the absence of any other information about the relative importance of those that did not respond compared with those that did, one very crude technique may be to multiply the reported data by 100/30 to derive an estimate for the total. However, such a crude technique should be a last resort, but it might be reasonable for a sector if all enterprises in it tend to be similar. If there is some indication of the size of the nonrespondents to the survey (such as based on other surveys filed by the same enterprises, relative employment sizes, value added, or asset size) that could be used to gain some measure of the relative importance of those that have not reported compared with those that have, then this information should be applied, even if it is only an overall indication of size. In addition, telephone contact should be made to obtain useful input from the enterprises that are considered to be among the largest nonrespondents to the survey, because it is extremely important to find some method of estimating the data for these enterprises as accurately as possible. Published information such as financial statements on the internet may be useful to impute data for large nonrespondent enterprises. If resources permit it can be useful to inform a large nonrespondent what imputed data for that enterprise will be included, indistinguishably, in the aggregate statistics. The enterprise might be inclined to provide better data.

2.62 Another option, where positions data are being surveyed and there has been a low response rate might be, in the absence of any other information, to use the comparative weights for accumulated transactions reported by those that reported transactions for balance of payments purposes versus those that have not and to apply the weight for estimating missing position data.

H. Editing/Validating Collected Data

2.63 If, despite all the preparations, survey respondents submit poor quality data, much good work will nonetheless have been accomplished, because the groundwork will have been laid for follow-on, more successful effort.

2.64 The compiler conducting a survey for the first time needs to be especially vigilant in checking data supplied. The more experience the respondent has in completing the survey form, up to a point, the less likely significant errors will be made.

2.65 The compiler can conduct editing/validation control checks on a number of levels. The compiler can include checks in the survey form, cross-check survey data against other reported data, and devise analytical checks.

2.66 This section provides some guidance on a range of possible checks. Nonetheless, it should be borne in mind that the more the survey form fits the domestic circumstances, and the more that survey respondents are consulted about what is required of them, then, inter alia, the greater the probability of receiving good quality data. Perhaps just as important, the more groundwork that is laid, the greater the likelihood of cooperation, if the national compiler needs to question the respondent about reported data.

Data editing/validation through the questionnaire form

2.67 The sooner errors are spotted and corrected, the better. For this reason, it is recommended that the compiler consider devising a survey form that explicitly includes quality control checks and/or requires extra information, which can be used as a consistency check. However, if extra information is required, it should be kept to the minimum necessary, and the compiler should be clear as to the purpose of the collection.

2.68 Among the tools that can help raise data quality are various computer systems that can allow for internal edits, including commercial spreadsheets; relational database management systems; and time series database management systems. Simple computer-checking procedures can be written to process reported data. Edit checks could include some of the following:

- Are the results consistent with what might be expected? (This question poses some difficulties for surveys conducted for the first time, but the compiler may be able to use other data sources such as those used for input to the national accounts as an indication of what might be expected from any given respondent.)

- If total assets and liabilities (including shareholders' funds) are to be reported, are they equal?

- Are reported transactions consistent with the reported positions data after taking account of such nontransaction changes as price and exchange rate changes, write-

offs, reclassifications? To use this particular quality control check, positions data are required for points in time.

- Where income is reported, do the rates of return on assets/liabilities make sense in light of rates of return available for other enterprises in the economy?

- More generally, the compiler could require an official of the reporting company to certify that the information provided is complete and accurate; this could help ensure data quality and promote timely reporting. Similarly, requesting the name of a contact person helps ensure that follow-up enquiries are efficiently directed.

Data editing/validation through analytical checks

2.69 The degree to which analytical checks can be developed depends on the availability of comparable data. If data are available, the following checks can be devised:

- If stock data and flow data are independently assembled, then an attempt could be made to reconcile these data. To do so requires consideration of effects caused by different prices, exchange rates, reporting thresholds, and other factors like write-offs. Checks can be applied at the individual respondent and aggregate data levels.

- For surveys where financial position data are being collected for the first time, but transactions data are available, position data could be compared to transaction data. If sizeable transactions data appear opposite a country for which reported positions are small, it should be investigated.

Advantages of a survey as a data source

2.70 The main advantage of using surveys as a data source may be that surveyors have more direct control over the information collected, because it is not a by-product of administrative or financial systems. In countries where questions can be added to regular surveys, further useful information – also for analytical reason – can be obtained whenever the survey is administered. Therefore a survey meets the requirements for compiling the balance of payments and IIP.

2.71 Surveys can also provide insights about for what purpose the transaction is used, which can be useful information for the compiler when evaluating the coverage of data obtained from other sources. It also may provide information to help estimate bilateral flows, which is very useful as transactions become increasingly complex.

2.72 Enterprise surveys of financial components can be conducted specifying opening and closing stocks reconciled with transactions, so providing in principle more consistent data and permitting sound consistency checks.

2.73 The reporters have greater knowledge of their transactions and are able to convey more accurate information regarding foreign counterparts and level of detail. Therefore enterprise surveys cause less misclassification.

2.74 Enterprise surveys are usually conducted by mail, email, web applications provided by the compiler, or personal interview; therefore the reporting burden and the costs for the reporter usually are moderate.

2.75 Information on other economic activity additional to balance of payments activities can be easily collected through surveys for analytical purposes and for quality control.

Disadvantages of a survey as a data source

2.76 There is the possibility of sampling error, particularly where the target population under study is relatively rare and the number of respondents in the sample small. The target population may not be uniformly distributed among the whole population. Special sampling techniques may be needed to identify them and include them in statistically representative samples.

2.77 Samples may not represent the desired target population. The behavior of the sample could differ from the whole population and the resulting estimates obtained in this way will contain a bias.

2.78 There is the possibility of non-sampling error. For example, the most significant error in information on personal transfers is that it may be underreported, because these data are often considered sensitive by respondents.

2.79 Conducting a survey may be costly. The costs for the compiler of using surveys vary greatly between countries, in line with the cost of enumeration and the cost of obtaining technical advice. Obtaining estimates with greater precision usually requires larger samples, which increases costs or decreases freshness. The costs of designing surveys vary according to the sampling method used. Adding questions to an existing survey may be a cheaper method of obtaining survey data, but resulting usable sample sizes may be small unless consecutive samples are pooled because some cases are likely to be relatively rare.

Box 2.2. Conducting a Private Capital Flows Survey in Anglophone Africa (the United Kingdom Department for International Development - DFID)

Background

This case study gives a brief overview of the IMF/DFID Enhanced Data Dissemination Initiative (EDDI) for Anglophone Africa to introduce annual private capital flows surveys (PCFS) in a number of African countries. The PCFS provides direct reporting of private sector cross-border financial flows and stocks that can be used to compile the balance of payments financial account and related IIP account. In some cases, the survey was expanded to include current and capital account transactions and to collect data to meet other local stakeholder needs. However, the central focus of the PCFS was on the balance of payments financial account and the IIP largely because data sources in this area were poor and because of a heightened interest in the IIP.

Identification of the survey population

The survey's frame was mainly constructed by including all listed companies and consulting with the main industry associations, large enterprise groups, and relevant government and public sector bodies (such as regulatory agencies, investment agencies, and export processing zones). The survey frames were mostly limited to private enterprises although some participating countries included selectively public corporations.

Approach applied in conducting the survey

After the identification of the survey population, an exploratory survey module was added by some countries to existing large frame enterprise surveys or censuses conducted for other statistical purposes. It helped expanding the frame to a size that would support a census of enterprises that met a threshold of foreign liabilities and assets. The result showed that all enterprises that met a threshold of foreign assets and liabilities were included in the PCFS. The compilers decided to conduct annual surveys with periodic larger frames and surveys with smaller frames for intermediate years not yet explored. The establishment of close working relationships by balance of payments and IIP compilers with industry associations, the local stock exchange for listed companies, and key major enterprises prompted the increased use of email communication and resulted in initiatives to establish the electronic reporting by some enterprises, although this is at an early stage of development.

For most of the countries participating in the EDDI project, a census of enterprises that met a threshold of cross-border financial flows and stocks was established. For others, further steps were required to finalize the survey frame through the use of exploratory surveys as well as more detailed checks of the ITRS data for large financial account transactions.

Data review process

Most countries built internal review checks into the questionnaire. Control mechanisms were used to check the reported data against previous survey responses, enterprises financial statements, and/or annual reports to shareholders. The latest enterprise financial statement and/or annual report to shareholders was requested from reporters together with the completed questionnaire.

During the project the internal review checks built into the questionnaire were adjusted. Investigations showed that the process of data review was best completed if the database was designed to facilitate the data review process. Some countries in the project successfully established

databases based on relevant data management software that provides tools for scrolling through an entire enterprise survey to review responses on an enterprise-by-enterprise basis, looking for gaps and outliers that escaped the attention of those completing the data entry screens.

Grossing up for non-response

Survey response was running at over 70 percent for most countries, although some major enterprises were reluctant to respond. Response had generally improved with successive surveys. Pre-survey sensitization meetings helped in this regard, as well as the publication of reports on the outcome of the surveys. Techniques for grossing up for non-response had evolved in the course of the project.

Because most cross-border financing is accounted for by the major enterprises, estimates for non-responding enterprises were normally made on an enterprise by enterprise basis for the larger enterprises using available data, such as enterprise financial statements, the application of carry forward techniques to completed questionnaires for the enterprise in earlier surveys, and the use of ITRS data for large financing transactions. For smaller enterprises, grossing up techniques were applied when an earlier large frame PCFS was conducted that allowed estimates for non-response to be made on the basis of the earlier survey. Some countries applied different grossing up techniques for large and small enterprises. For example, it is likely that for small locally owned enterprises portfolio investment is small and borrowing from abroad is likely to comprise trade credit received. This information was taken into account when decided how to gross up the resulting data.

The need for grossing up also arose when successive enterprise surveys had different survey frames, or when there were differential survey responses. Earlier experience with enterprise surveys in the region had been to conduct each survey on a standalone basis with its own database. This was especially where spreadsheet software was used to aggregate the reported data.

Databases were developed using appropriate data management software where it was possible to bring the results of successive enterprise surveys into a single database. These databases made it possible to track the data reported by a single enterprise across all years that had been reported, which may be particularly important where there were mergers and acquisitions.

Other examples of grossing up (or down) techniques that were used concern the use of alternative data sources for balance of payments and IIP compilation (such as ITRS or administrative data sources for investment approvals) that cover earlier/later periods. In the case of trade credit, if a relationship is found between survey data and customs or ITRS to obtain data for imports and exports of goods, some countries worked plans to apply grossing up techniques to estimate trade credit for periods not covered by enterprise surveys.

Incorporation of survey results in balance of payments and IIP statistics

On data dissemination issues, the EDDI project follows the IMF's Data Quality Assessment Framework (DQAF) for balance of payments statistics. Since the project is expected to result in substantial revision of balance of payments to ensure their consistency with directly reported IIP statistics, all participating countries were expected to follow the DQAF guidelines on revision policy. These required that the revised data were announced by a press release that explained the reasons for the revisions being made to previously published statistics together with documentation of the details of the revisions that have resulted, and plans for the future. Publication of the resulting balance of payments and IIP data was assessed against the DQAF standards for methodological soundness (in conformity with the *BPM6,* accuracy and reliability, serviceability (periodicity, timeliness, and

consistency), and accessibility.

Encountered difficulties

- For some countries, tax records proved to be of limited use for the identification of the survey population as they did not provide a useful tool for identifying enterprises with significant foreign liabilities and assets.
- In most cases where no foreign funding was involved, budgetary considerations limited the size of the survey frame. In those cases, it was expected that the results could be recast using appropriate grossing up techniques. A larger frame survey should be conducted at a later stage.
- Experience in the region in using mail, email, or telephone for contacting enterprises or sending out enterprise questionnaires was disappointing. For the most part, the use of email and other electronic media had only proven effective once a working relationship with a counterpart in the enterprise had been established, and even then was best suited for collecting data from the larger enterprises. In practice, for most countries in the region, the preferred way of delivering enterprise questionnaires and collecting completed questionnaires was through the use of enumerators, particularly when the enumerator presented questionnaires to the enterprise physically and returned after two weeks to collect completed questionnaires.

Lessons learned

- Assuring that all large enterprises were captured in the survey frame was often difficult. In some cases, ITRS data for large cross-border financial transactions that were classified by type of instrument and by the name of the accountholder had proven useful as a check on whether all large enterprises, especially those engaged in major investment projects, were included.
- Pre-survey sensitization meetings were very important to increase the response rate.
- For some countries the best way to collect data for the first time was by sending enumerators to the enterprises instead of mail, email or telephone surveys.
- Well defined databases encompassing the enterprise survey data were important in increasing the quality of balance of payments and IIP statistics.