Chapter 9. Methods for collecting data on business innovation

This chapter provides guidance on methodologies for collecting data on business innovation, based on the concepts and definitions introduced in previous chapters. The guidance is aimed at producers of statistical data on innovation as well as advanced users who need to understand how innovation data are produced. While acknowledging other potential sources, this chapter focuses on the use of business innovation surveys to collect data on different dimensions of innovation-related activities and outcomes within the firm, along with other contextual information. The guidance within this chapter covers the full life cycle of data collection, including setting the objectives and scope of business innovation surveys; identifying the target population; questionnaire design; sampling procedures; data collection methods and survey protocols; post-survey data processing, and dissemination of statistical outputs.

9.1. Introduction

9.1.   This chapter provides guidance on methodologies for collecting data on business innovation. As noted in Chapter 2, methodological guidance for the collection of data on innovation is an essential part of the measurement framework for innovation. Data on innovation can be obtained through object-based methods such as new product announcements on line or in trade journals (Kleinknecht, Reijnen and Smits, 1993), and from expert assessments of innovations (Harris, 1988). Other sources of innovation data include annual corporate reports, websites, social surveys of employee educational achievement, reports to regional, national and supranational organisations that fund research and experimental development (R&D) or innovation, reports to organisations that give out innovation prizes, university knowledge transfer offices that collect data on contract research funded by firms and the licensing of university intellectual property, business registers, administrative sources, and surveys of entrepreneurship, R&D and information and communication technology (ICT) use. Many of these existing and potential future sources may have "big data" attributes, namely they are too large or complex to be handled by conventional tools and techniques.

9.2.   Although useful for different purposes, these data sources all have limitations. Many do not provide representative coverage of innovation at either the industry or national level because the data are based on self-selection: only firms that choose to make a product announcement, apply for R&D funding, or license knowledge from universities are included. Information from business registers and social, entrepreneurship, and R&D surveys is often incomplete, covering only one facet of innovation. Corporate annual reports and websites are inconsistent in their coverage of innovation activities, although web-scraping techniques can automate searches for innovation activities on documents posted on line and may be an increasingly valuable source of innovation data in the future. Two additional limitations are that none of these sources provide consistent, comparable data on the full range of innovation strategies and activities undertaken by all firms, as discussed in Chapters 3 to 8, and many of these sources cannot be accurately linked to other sources. Currently, the only source for a complete set of consistent and linkable data is a dedicated innovation survey based on a business register.

9.3.   The goal of a business innovation survey is to obtain high-quality data on innovation within firms from authoritative respondents such as the chief executive officer or senior managers. A variety of factors influence the attainment of this goal, including coverage of the target population, the frequency of data collection, question and questionnaire design and testing, the construction of the survey sample frame, the methods used to implement the survey (including the identification of an appropriate respondent within the surveyed unit) and post-survey data processing. All of these topics are relevant to national statistical organisations (NSOs) and to international organisations and researchers with an interest in collecting data on innovation activities through surveys and analysing them.

9.4.   Business innovation surveys that are conducted by NSOs within the framework of national business statistics must follow national practices in questionnaire and survey design. The recommendations in this chapter cover best practices that should be attainable by most NSOs. Surveys implemented outside of official statistical frameworks, such as by international organisations or academics, will benefit from following the recommendations in this chapter (OECD, 2015a). However, resource and legal restrictions can make it difficult for organisations to implement all best practices.

9.5.   The decision on the types of data to collect in a survey should be taken in consultation with data users, including policy analysts, business managers and consultants,

academics, and others. The main users of surveys conducted by NSOs are policy makers and policy analysts and consequently the choice of questions should be made after consultations with those government departments and agencies responsible for innovation and business development. Surveys developed by academics could also benefit from consultations with governments or businesses.

9.6.   The purpose(s) of data collection, for instance to construct national or regional indicators or for use in research, will largely influence survey methodology choices. The sample can be smaller if only indicators at the national level are required, whereas a larger sample is necessary if users require data on sub-populations, longitudinal panel data, or data on rare innovation phenomena. In addition, the purpose of the survey will have a strong influence on the types of questions to be included in the survey questionnaire.

9.7.   This manual contains more suggestions for questions on innovation than can be included in a single survey. Chapters 3 to 8 and Chapter 10 recommend key questions for collection on a regular basis and supplementary questions for inclusion on an occasional basis within innovation survey questionnaires. Occasional questions based on the supplementary recommendations or on other sections of the manual can be included in one-off modules that focus on specific topics or in separate, specialised surveys. The recommendations in this chapter are relevant to full innovation surveys, specialised surveys, and to innovation modules included in other surveys.

9.8.   This chapter provides more details on best practice survey methods than previous editions of this manual. Many readers from NSOs will be familiar with these practices and do not require detailed guidance on a range of issues. However, this edition is designed to serve NSOs and other producers and users of innovation data globally. Readers from some of these organisations may therefore find the details in this chapter of value to their work. In addition to this chapter, other sources of generic guidelines for business surveys include Willeboordse (ed.) (1997) and Snijkers et al. (eds.) (2013). Complementary material to this manual's online edition will provide relevant links to current and recent survey practices and examples of experiments with new methods for data collection (http://oe.cd/oslomanual).

9.9.   The chapter is structured as follows: Section 9.2 covers the target population and other basic characteristics of relevance to innovation surveys. Questionnaire and question design are discussed in section 9.3. A number of survey methodology issues are discussed in the subsequent sections including sampling (section 9.4), data collection methods (section 9.5), survey protocol (section 9.6) and post-survey processing (section 9.7). The chapter concludes with a brief review of issues regarding the publication and dissemination of results from innovation surveys (section 9.8).

9.2. Population and other basic characteristics for a survey

9.2.1. Target population

9.10.   The Business enterprise sector, defined in Chapter 2 and OECD (2015b), is the target for surveys of business innovation. It comprises:

- All resident corporations, including legally incorporated enterprises, regardless of the residence of their shareholders. This includes quasi-corporations, i.e. units capable of generating a profit or other financial gain for their owners, recognised by law as separate legal entities from their owners, and set up for the purpose of engaging in market production at prices that are economically significant. They include both financial and non-financial corporations.

- The unincorporated branches of non-resident enterprises deemed to be resident and part of this sector because they are engaged in production on the economic territory on a long-term basis.

- All resident non-profit institutions that are market producers of goods or services or serve businesses. This includes independent research institutes, clinics, and other institutions whose main activity is the production of goods and services for sale at prices designed to recover their full economic costs. It also includes entities controlled by business associations and financed by contributions and subscriptions.

9.11.   The Business enterprise sector includes both private enterprises (either publicly listed and traded, or not) and government-controlled enterprises (referred to as "public enterprises" or "public corporations"). For public enterprises, the borderline between the Business enterprise and Government sectors is defined by the extent to which the unit operates on a market basis. If a unit's principal activity is the production of goods or services at economically significant prices it is considered to be a business enterprise.

9.12.   Consistent with the definition in the System of National Accounts (SNA) (EC et al., 2009), the residence of each unit is the economic territory with which it has the strongest connection and in which it engages for one year or more in economic activities. An economic territory can be any geographic area or jurisdiction for which statistics are required, for instance a country, state or province, or region. Businesses are expected to have a centre of economic interest in the country in which they are legally constituted and registered. They can be resident in different countries than their shareholders and subsidiary firms may be resident in different countries than their parent organisations.

9.13.   The main characteristics of the target population that need to be considered for constructing a sample or census are the type of statistical unit, the unit's industry of main activity, the unit's size, and the geographical location of the unit.

9.2.2. Statistical units and reporting units

9.14.   Firms organise their innovation activities at various levels in order to meet their objectives. Strategic decisions concerning the financing and direction of innovation efforts are often taken at the enterprise level. However, these decisions can also be taken at the enterprise group level, regardless of national boundaries. It is also possible for managers below the level of the enterprise (i.e. establishment or kind-of-activity unit [KAU]) to take day-to-day decisions of relevance to innovation.

9.15.   These decisions can cut across national borders, especially in the case of multinational enterprises (MNEs). This can make it difficult to identify and survey those responsible for decision-making, particularly when NSOs or other data collection agencies only have the authority to collect information from domestic units.

Statistical unit

9.16.   A **statistical unit** is an entity *about which* information is sought and for which statistics are ultimately compiled; in other words, it is the institutional unit of interest for the intended purpose of collecting innovation statistics. A statistical unit can be an **observation unit** for which information is received and statistics are compiled, or an **analytical unit** which is created by splitting or combining observation units with the help of estimations or imputations in order to supply more detailed or homogeneous data than would otherwise be possible (UN, 2007; OECD, 2015b).

9.17.   The need to delineate statistical units arises in the case of large and complex economic entities that are active in different industry classes, or have units located in different geographical areas. There are several types of statistical units according to their ownership, control linkages, homogeneity of economic activity, and their location, namely enterprise groups, enterprises, establishments (a unit in a single location with a single productive activity), and KAUs (part of a unit that engages in only one kind of productive activity) (see OECD [2015b: Box 3.1] for more details). The choice of the statistical unit and the methodology used to collect data are strongly influenced by the purpose of innovation statistics, the existence of records of innovation activity within the unit, and the ability of respondents to provide the information of interest.

9.18.   The statistical unit in business surveys is generally the **enterprise**, defined in the SNA as the smallest combination of legal units with "autonomy in respect of financial and investment decision-making, as well as authority and responsibility for allocating resources for the production of goods and services" (EC et al., 2009; OECD, 2015b: Box 3.1).

9.19.   Descriptive identification variables should be obtained for all statistical units in the target population for a business innovation survey. These variables are usually available from statistical business registers and include, for each statistical unit, an identification code, the geographic location, the kind of economic activity undertaken, and the unit size. Additional information on the economic or legal organisation of a statistical unit, as well as its ownership and public or private status, can help to make the survey process more effective and efficient.

Reporting units

9.20.   The **reporting unit** (i.e. the "level" within the business *from which* the required data are collected) will vary from country to country (and potentially within a country), depending on institutional structures, the legal framework for data collection, traditions, national priorities, survey resources and ad hoc agreements with the business enterprises surveyed. As such, the reporting unit may differ from the required statistical unit. It may be necessary to combine, split, or complement (using interpolation or estimation) the information provided by reporting units to align with the desired statistical unit.

9.21.   Corporations can be made up of multiple establishments and enterprises, but for many small and medium-sized enterprises (SMEs) the establishment and the enterprise are usually identical. For enterprises with heterogeneous economic activities, it may be necessary for regional policy interests to collect data for KAUs, or for establishments. However, sampling establishments or KAUs requires careful attention to prevent double counting during data aggregation.

9.22.   When information is only available at higher levels of aggregation such as the enterprise group, NSOs may need to engage with these units to obtain disaggregated data, for instance by requesting information by jurisdiction and economic activity. This will allow better interoperability with other economic statistics.

9.23.   The enterprise group can play a prominent role as a reporting unit if questionnaires are completed or responses approved by a central administrative office. In the case of holding companies, a number of different approaches can be used, for example, asking the holding company to report on the innovation activities of enterprises in specific industries, or forwarding the questionnaire, or relevant sections, to other parts of the company.

9.24.   Although policy interests or practical considerations may require innovation data at the level of establishments, KAUs, and enterprise groups, it is recommended, wherever

possible, to collect data at the **enterprise level** to permit international comparisons. When this is not possible, careful attention is required when collecting and reporting data on innovation activities and expenditures, as well as linkage-related information, that may not be additive at different levels of aggregation, especially in the case of MNEs. Furthermore, innovation activities can be part of complex global value chains that involve dispersed suppliers and production processes for goods and services, often located in different countries. Therefore, it is important to correctly identify whenever possible statistical units active in global value chains (see Chapter 7) in order to improve compatibility with other data sources (such as foreign investment and trade surveys).

Main economic activity

9.25.   Enterprises should be classified according to their main economic activity using the most recent edition of the United Nation's (UN) International Standard Industrial Classification (ISIC Rev.4) (see UN, 2008) or equivalent regional/national classifications. ISIC supports international comparability by classifying industries into economic activities by section, division, group and class, though in most cases the target population can be defined using the section and division levels. The recommendations given below use the sections and divisions as defined in ISIC Rev.4. These should be updated with future revisions of ISIC.

9.26.   When there is significant uncertainty about the true economic activity of firms (for instance if this information is not available from a business register, refers to non-official classifications or is likely to be out of date) innovation surveys can include a question on the main product lines produced by each firm and, if possible, questions on the relative importance of different types of product lines (for instance the contribution of different product categories to turnover). This information is required to assign an economic activity to the enterprise, both for stratification, sampling and analytical purposes.

9.27.   As noted in Chapters 1 and 2, this manual recommends the collection of innovation data for businesses in most ISIC-defined industries, with some qualifying exceptions discussed below. Key considerations when defining the recommended scope of business innovation surveys by economic activity, especially for international comparison purposes, are the prevalence of non-business actors in an industry, the presence of specific measurement challenges such as unstable business registers, and previous international experience measuring innovation within an industry.

9.28.   Table 9.1 provides the broad structure of industries by ISIC Rev.4 at the section and division level and identifies economic activities that are recommended for international comparisons, supplementary economic activities that may be worth including for national purposes, and economic activities that are not currently recommended for surveys of innovation in the Business sector.

9.29.   The **recommended** economic activities for national data collection and for international comparisons include ISIC Rev.4 sections B to M inclusive with the exception of section I (Accommodation and food service activities). In these areas there is substantive national and international comparative experience with data collection.

9.30.   **Supplementary** economic activities that are worth collecting, but are still largely untested from an international comparison perspective, include ISIC Rev.4 sections A (Agriculture, forestry and fishing), I (Accommodation and food services activities), N (Administrative and support activities), and divisions 95-96 of section S (Repair activities, other personal service activities). For these industries, the international standardisation of business registers is still incomplete (particularly for agriculture) and current experience is

limited to surveys in only a few countries. Any ongoing efforts should provide better guidance for innovation measurement in the future.

Table 9.1. Economic activities for inclusion in international comparisons of business innovation

Based on UN ISIC Rev.4 sections and divisions

Section	Division	Description
Economic activities recommended for inclusion for international comparisons		
B	05-09	Mining and quarrying
C	10-33	Manufacturing
D	35	Electricity, gas, steam and air conditioning supply
E	36-39	Water supply; sewerage, waste management and remediation activities
F	41-43	Construction
G	45-47	Wholesale and retail trade; repair of motor vehicles and motorcycles
H	49-53	Transportation and storage
J	58-63	Information and communication
K	64-66	Financial and insurance activities
L	68	Real estate activities
M	69-75	Professional, scientific and technical activities
Supplementary economic activities for national data collections		
A	01-03	Agriculture, forestry and fishing
I	55-56	Accommodation and food service activities
N	77-82	Administrative and support service activities
S	95-96	Repair activities, other personal service activities
Economic activities not recommended for data collection		
O	84	Public administration and defence; compulsory social security
P	85	Education
Q	86-88	Human health and social work activities
R	90-93	Arts, entertainment and recreation
S	94	Membership organisations
Economic activities outside the scope of this manual		
T	97-98	Activities of households as employers; activities of households for own-use
U	99	Activities of extraterritorial organisations and bodies

9.31.   A number of economic activities are *not generally recommended* for data collection by business innovation surveys and should be excluded from international comparisons of business innovation. From an international comparison perspective, sections O (Public administration), P (Education), Q (Human health and social work), R (Arts, entertainment and recreation) and division 94 of section S (Membership organisations) are not recommended for inclusion because of the dominant or large role of government or private non-profit institutions serving households in the provision of these services in many countries. However, there may be domestic policy demands for extending the coverage of national surveys to firms active in these areas, for example if a significant proportion of units active in this area in the country are business enterprises, or if such firms are entitled to receive public support for their innovation activities.

9.32.   Other sections recommended for exclusion are dominated by actors engaged in non-market activities and therefore *outside the scope* of this manual, namely section T (Households) and section U (Extraterritorial bodies).

Unit size

9.33.   Although innovation activity is generally more extensive and more frequently reported by larger firms, units of all sizes have the potential to be innovation-active and should be part of the scope of business innovation surveys. However, smaller business units, particularly those with higher degree of informality (e.g. not incorporated as companies, exempt from or not declaring some taxes, etc.), are more likely to be missing from statistical business registers. The relative importance of such units can be higher in countries in earlier stages of development. Comparing data for countries with different types of registers for small firms and with varying degrees of output being generated in the informal economy can therefore present challenges. An additional challenge, noted in Chapter 3, stems from adequately interpreting innovation data for recently created firms, for which a substantial number of activities can be deemed to be new to the firm.

9.34.   Therefore, for international comparisons, it is recommended to limit the scope of the target population to comprise all statistical business units with *ten or more persons employed* and to use *average headcounts* for size categories. Depending on user interest and resources, surveys can also include units with fewer than ten persons employed, particularly in high technology and knowledge-intensive service industries. This group is likely to include start-ups and spin-offs of considerable policy interest (see Chapter 3).

9.2.3. Data linkage

9.35.   An official business register is often used by NSOs to identify the sample for the innovation survey and for the R&D, ICT, and general business statistics surveys. This creates opportunities for linking the innovation survey with other surveys in order to obtain interval data on several variables of interest, such as on R&D, ICT, employment, turnover, exports or investments. Over the years, an increasing number of NSOs have used data linkage to partially eliminate the need to collect some types of data in the innovation survey, although data linkage is only possible when the surveys to be linked use the same statistical units, which for NSOs is usually the enterprise.

9.36.   Data linkage can reduce respondent burden, resulting in higher response rates, and improve the quality of interval data that are obtained from mandatory R&D and business surveys. However, questions must be replicated in an innovation survey when respondents need a reference point for related questions, either to jog their memory or to provide a reference for calculating subcategories or shares. For example, questions on innovation expenditures should include a question on R&D expenditures for reference, and questions on the number (or share) of employees with different levels of educational attainment should follow a question on the total number of employees. Once the survey is completed, the innovation survey values for R&D, employment, or other variables can be replaced for some analyses by values from the R&D and business surveys, if analysis indicates that this will improve accuracy.

9.37.   Another option created by the ability to combine administrative and survey data is to pre-fill online innovation questionnaires with data obtained from other sources on turnover, employment, R&D expenditures, patent applications, etc. These can provide immediate reference points for respondents and reduce response burden. A disadvantage is that pre-filled data could be out of date, although older data could still be useful for pre-filling data for the first year of the observation period. Respondents should also be given an option to correct errors in pre-filled data.

9.38.   Linkage to structural business statistics data on economic variables after a suitable time lag (one or more years after the innovation survey) is useful for research to infer causal

relationships between innovation activities and outcomes. Relevant outcomes include changes in productivity, employment, exports and revenue.

9.39.   Selected innovation questions may be added occasionally to other surveys to assist in improving, updating and maintaining the innovation survey frame.

9.2.4. Frequency of data collection

9.40.   The frequency of innovation surveys depends on practical considerations and user needs at the international, national and regional level. Considerations such as cost, the slow rate of change in many variables, the effect of frequent surveys on response burden, and problems due to overlapping observation periods between consecutive surveys influence the recommended frequency for innovation surveys. The importance of innovation for economic growth and well-being creates a policy demand for more frequent and up-to-date data collected on an annual basis, particularly for innovation activities that can change quickly. Annual panel surveys can also facilitate analysis of the lag structure between innovation inputs and outputs, or the effects of innovation on economic performance (see Chapter 11).

9.41.   It is recommended to conduct innovation surveys every **one to three years**. For a frequency of two or three years, a shorter survey that only collects key innovation variables can be conducted in alternating years, resources permitting. However, caution is required when comparing the results of short and long surveys because responses can be affected by survey length (see section 9.3 below). Information on innovation can also be obtained from the Internet or other sources in years without an innovation survey. Options for using alternative sources of innovation data in non-survey years have yet to be investigated in detail.

9.2.5. Observation and reference periods

9.42.   To ensure comparability among respondents, surveys must specify an observation period for questions on innovation. The observation period is the length of time covered by a majority of questions in a survey. In order to minimise recall bias, it is recommended that the observation period **should not exceed three years**. The reference period is the final year of the overall survey observation period and is used as the effective observation period for collecting interval level data items, such as expenditures or the number of employed persons. The reference and observation periods are identical in surveys that use a one-year observation period.

9.43.   The length of the observation period qualifies the definition of innovation and therefore the share of units that are reported as innovative (see Chapter 3). For example, the choice of an observation period can affect comparisons between groups of units (e.g. industries) that produce goods or services with varying life cycles (industries with short product life cycles are more likely to introduce product innovations more frequently). This has implications for interpretability and raises the need for adequate standardisation across national surveys (see Chapter 11).

9.44.   In some instances, interpretation issues favour a longer observation period. For example, if an innovation project runs over several years, a short observation period might result in assigning different innovation activities and outputs to different years, such as the use of co-operation, the receipt of public funding, and sales from new products. This could hamper some relevant analyses on innovation patterns and impacts.

9.45.   Data quality concerns favour a shorter observation period in order to reduce recall errors. This applies for instance when respondents forget to report an event, or from telescoping errors when respondents falsely remember events that happened before the observation period as occurring during that period.

9.46.   The quality advantages of short observation periods and the potential interpretation advantages of longer observation periods may be combined through the construction of a longitudinal panel linking firms in consecutive cross-sectional innovation surveys (see subsection 9.4.3 below). For example, if the underlying data have a one-year observation period, the innovation status of firms over a two- (three-) year period can be effectively calculated from data for firms with observations over two (or three) consecutive annual observation periods. Additional assumptions and efforts would be required to deal with instances where repeated observations are not available for all firms in the sample, for example due to attrition, or the use of sampling methods to reduce the burden on some types of respondents (e.g. SMEs). A strong argument in favour of a longitudinal panel survey design is that it enhances the range of possible analyses of causal relationships between innovation activities and outcomes (see subsection 9.4.3 below).

9.47.   Observation periods that are longer than the frequency of data collection can affect comparisons of results from consecutive surveys. In such cases, it can be difficult to determine if changes in results over time are mainly due to innovation activities in the non-overlapping period or if they are influenced by activities in the period of overlap with the previous survey. Spurious serial correlation could therefore be introduced as a result.

9.48.   At the time of publication of this manual, the observation period used by countries varies between one and three years. This reduces international comparability for key indicators such as the incidence of innovation and the rate of collaboration with other actors. Although there is currently no consensus on what should be the optimal length of the generic observation period (other than a three-year maximum limit), convergence towards a common observation period would considerably improve international comparability. It is therefore recommended to conduct, through concerted efforts, additional experimentation on the effects of different lengths for the observation period and the use of panel data to address interpretation issues. The results of these experiments would assist efforts to reach international agreement on the most appropriate length for the observation period.

9.3. Question and questionnaire design

9.49.   Chapters 3 to 8 of this manual identify different concepts and characteristics of business innovation for measurement. These need to be operationalised in the form of questions that create useful data for empirical analysis, as indicated in Figure 9.1.

9.50.   The operationalisation of theoretical concepts can result in a number of possible errors that can be minimised through careful question and questionnaire design. This manual does not provide full examples of survey questions because the phrasing of final questions is likely to differ depending on contextual conditions that vary across and within countries. Instead, the following sections provide guidelines for best practice in the design of questions and the layout of the questionnaire. Good question design and questionnaire layout can improve data quality, increase response rates, and improve comparability across different survey methods (see subsection 9.5.4 below).

9.51.   The design of individual questions and questionnaire layout are influenced by the ability to obtain data from other sources (which makes some questions unnecessary), and by the choice of survey method (see section 9.5 below). For example, grid or matrix questions are difficult and time-consuming when heard, as in telephone surveys, but are easily understood when presented visually, as in online and mailed questionnaire surveys. These differences in survey methods need to be taken into consideration when multiple methods are used.

Figure 9.1. From innovation theory to innovation data
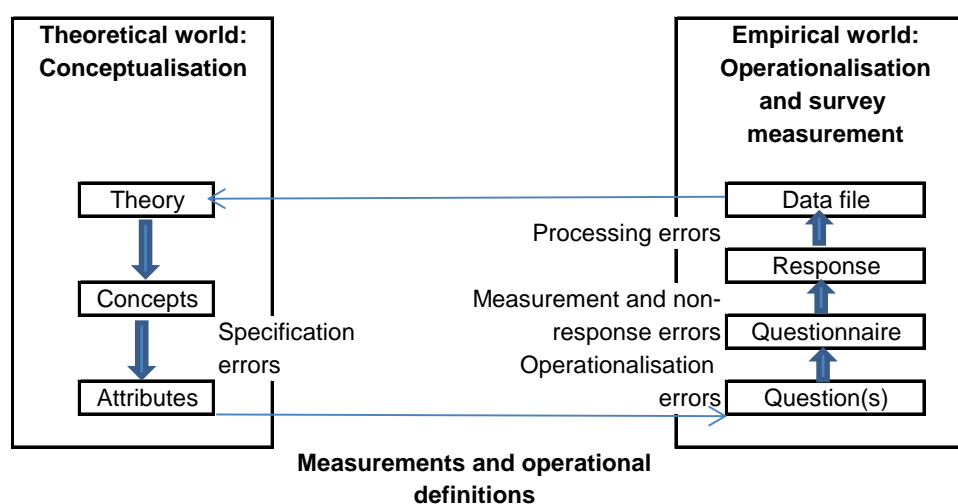
Source: Based on Snijkers and Willimack (2011), "The missing link: From concepts to questions in economic surveys".

9.3.1. Question design

Question adaptation and translation

9.52.   All questions need to be carefully adapted and translated to ensure that respondents interpret the questions as intended by the concepts and definitions in this manual. First, many concepts and definitions cannot be literally applied as questions. For example, more than one question could be required to obtain data that capture a single concept (see Chapter 3). Second, key terms must be adapted to match the language used by respondents in different cultural, regional and national contexts (Harkness et al. [eds.], 2010). For instance, Australia and Canada use the term "business" instead of "enterprise" because the latter is not part of common English usage in either country and is therefore open to misunderstanding. The words "enterprise" or "business" could also be confusing to respondents from establishments or KAUs.

9.53.   Translation issues are particularly important for innovation surveys that cover multiple countries or countries with more than one language, since even minor differences between national questionnaires can reduce the comparability of the results. These differences can stem from translation, changes in the order of questions, or from adding or deleting categories or questions. Translation needs to take account of country-specific circumstances (such as a country's legal system and regulations) to avoid misunderstandings of concepts and definitions.

Question comprehension and quality

9.54.   Questions need to be short, written in simple language, and unambiguous. It is important to eliminate repetition, such as when two questions ask for similar information, and to eliminate questions that ask for two or more information items (often identifiable by the use of "and" between two clauses). Wherever possible, concepts and definitions should be included in the questions because respondents often do not read supplementary information. The inclusion of explanatory information in footnotes or online hypertext links should be used as little as possible.

9.55.   Data quality can be improved by reducing respondent fatigue and by maintaining a motivation to provide good answers. Both fatigue and motivation are influenced by question length, but motivation can be improved by questions that are relevant and interesting to the respondent. The latter is particularly important for respondents from non-innovative units, who need to find the questionnaire relevant and of interest, otherwise, they are less likely to respond. Therefore, all questions should ideally be relevant to all units in all industries (Tourangeau, Rips and Rasinski, 2000).

9.56.   "Satisficing" refers to respondent behaviours to reduce the time and effort required to complete an online or printed questionnaire. These include abandoning the survey before it is completed (premature termination), skipping questions, non-differentiation (when respondents give the identical response category to all sub-questions in a question, for example answering "slightly important" to all sub-questions in a grid question), and speeding through the questionnaire (Barge and Gelbach, 2012; Downes-Le Guin et al., 2012). The main strategies for minimising satisficing are to ensure that the questions are of interest to all respondents and to minimise the length of the questionnaire. Non-differentiation can be reduced by limiting the number of sub-questions in a grid to no more than seven (Couper et al., 2013). Grid questions with more than seven sub-questions can be split into several subgroups. For instance, a grid question with ten sub-questions could be organised around one theme with six sub-questions, and a second theme with four.

Nominal and ordinal response categories

9.57.   Qualitative questions can use nominal response categories ("yes or no") or ordinal ones such as an importance or frequency scale. Nominal response categories are simple and reliable, but provide only limited information; while ordinal response categories may introduce a degree of subjectivity. Both types of questions may require a "don't know" or "not applicable" response category.

9.3.2. Questionnaire design

9.58.   The questionnaire should be as short as possible, logically structured, and have clear instructions. In voluntary surveys, unit response rates (the percentage of the sample that completes the questionnaire) decline with questionnaire length. The quality of responses can also decline for questions placed towards the end of a long questionnaire (Galesic and Bosnjak, 2009). Survey implementation factors that affect unit response rates are discussed in section 9.6 below.

9.59.   Respondent comprehension and willingness to respond can be affected by the questionnaire layout, with best practices being similar between printed and online questionnaires. Skip routines or branching instructions on printed questionnaires need to be clearly visible. The layout needs to incorporate national preferences for font sizes and the amount of blank space on a page. Instructions should be repeated wherever applicable to improve the likelihood that respondents follow them.

Filters

9.60.   Filters and skip instructions direct respondents to different parts of a questionnaire, depending on their answers to the filter questions. Filters can be helpful for reducing response burden, particularly in complex questionnaires. Conversely, filters can encourage satisficing behaviour whereby respondents answer "no" to a filter question to avoid completing additional questions.

9.61.   The need for filters and skip instructions can be minimised, for instance by designing questions that can be answered by all units, regardless of their innovation status. This can provide additional information of value to policy and to data analysis. However, filters are necessary in some situations, such as when a series of questions are only relevant to respondents that report one or more product innovations.

9.62.   The online format permits automatic skips as a result of a filter, raising concerns that respondents who reply to an online questionnaire could provide different results from those replying to a printed version which allows them to see skipped questions and change their mind if they decide that those skipped questions are relevant. When both online and printed questionnaires are used, the online version can use "greying" for skipped questions so that the questions are visible to respondents. This could improve comparability with the printed version. If paradata – i.e. the data about the process by which surveys are filled in – are collected in an online survey (see section 9.5 below), each respondent's path through the questionnaire can be evaluated to determine if greying has any effect on behaviour, for instance if respondents backtrack to change an earlier response.

Question order

9.63.   A respondent's understanding of a question can be influenced by information obtained from questions placed earlier in the questionnaire. Adding or deleting a question can therefore influence subsequent answers and reduce comparability with previous surveys or with surveys conducted in other jurisdictions.

9.64.   Questions on activities that are relevant to all units regardless of their innovation status should be placed before questions on innovation and exclude references to innovation. This applies to potential questions on business capabilities (see Chapter 5).

9.65.   Wherever possible, questions should be arranged by theme so that questions on a similar topic are grouped together. For instance, questions on knowledge sourcing activities and collaboration for innovation should be co-located. Questions on the contribution of external actors to a specific type of innovation (product or business process) need to be located in the section relating to that type of innovation.

9.3.3. Short-form questionnaires

9.66.   For many small units and units in sectors with little innovation activity, the response burden for a full innovation questionnaire can be high relative to their innovation activity, thereby reducing response rates. In such cases, shorter survey questionnaires that focus on a set of core questions could be useful. Short-form questionnaires can also be used to survey units that have not reported innovation activity in previous innovation surveys. However, empirical research for Belgium (Hoskens et al., 2016) and various developing countries (Cirera and Muzi, 2016) finds significant differences in the share of innovative firms among respondents to short and long questionnaires, with higher rates of innovative firms found for short questionnaires. These results suggest that comparisons of data on innovation from short-form and long-form questionnaires may reflect design factors that should be carefully taken into account.

9.3.4. Combining innovation and other business surveys

9.67.   Several NSOs have combined their innovation surveys with other business surveys, in particular with R&D surveys, due to the conceptual and empirical proximity between R&D and innovation. In principle, various types of business surveys can be integrated with

innovation surveys, for instance by combining questions on innovation with questions on business characteristics, ICT or knowledge management practices.

9.68. There are several advantages of combining surveys including:

- A combined survey can reduce the response burden for reporting units as long as the combined survey is shorter in length and less difficult than separate surveys due to avoiding repeated questions.

- A combined survey permits analyses of the relationship between innovation and other activities within the responding unit, for instance ICT use. This is advantageous if separate surveys cannot be linked or if the innovation survey and other surveys use different samples.

- A combined survey can reduce the printing and postage costs for questionnaires provided by mail and the follow-up costs for all types of surveys.

9.69. On the other hand, there are also disadvantages from combining surveys such as:

- Both the unit and item response rates can decline if the combined questionnaire is much longer than the length of the separate survey questionnaires. This is most likely to be a problem for voluntary surveys.

- If the topics are sufficiently distinct and relate to different functional areas within the business, several persons within an organisation, especially large ones, may need to be contacted in order to answer all questions.

- Combining an innovation and an R&D survey can result in errors in the interpretation of questions on innovation and R&D. Some respondents from units that do not perform R&D could incorrectly assume that innovation requires R&D, or that they are only being invited to report innovations based on R&D. This can lower the observed incidence of innovation, as reported in some countries that have experimented with combined R&D and innovation surveys (the observed incidence of R&D is not affected) (e.g. Wilhelmsen, 2012). In addition, some respondents could erroneously report innovation expenditures as R&D expenditures.

- The sampling frames for the innovation survey and other business surveys can differ. In the case of combining innovation and R&D surveys, the sample for innovation can include industries (and small units) that are not usually included in R&D surveys.

9.70. Based on the above considerations, the guidelines for combining an innovation survey with one or more other business surveys are as follows:

- A combined R&D and innovation survey needs to reduce the risk of conceptual confusion by non-R&D-performing units by using two distinct sections and by placing the innovation section first.

- Separate sections need to be used when combining innovation questions with other types of questions, such as on ICT or business characteristics. Questions that are relevant to all units should be placed before questions on innovation.

- A combined R&D and innovation survey can further reduce conceptual problems by ensuring that the R&D section is only sent to units that are likely to perform R&D.

- To avoid a decline in response rates, the length of a combined survey should be comparable to the summed length of the separate surveys, particularly for voluntary surveys.

- Care should be taken in comparisons of innovation results from combined surveys with the results from separate innovation surveys. Full details on the survey method should also be reported, including steps to reduce conceptual confusion.

9.71.   Therefore, as a general rule, this manual recommends **not combining R&D and innovation surveys** because of the drawbacks mentioned earlier, for instance by suggesting to some respondents that innovation requires R&D. Although untested, it seems at this point that there could be fewer problems with combining an innovation survey with other types of surveys, such as surveys on business strategy or business characteristics.

9.3.5. Questionnaire testing

9.72.   Innovation surveys undergo regular updates to adapt to known challenges and address emerging user needs. It is strongly recommended to subject all new questions and questionnaire layout features to cognitive testing in face-to-face interviews with respondents drawn from the survey target population.

9.73.   Cognitive testing, developed by psychologists and survey researchers, collects verbal information on survey responses. It is used to evaluate the ability of a question (or group of questions) to measure constructs as intended by the researcher and if respondents can provide reasonably accurate responses. The evidence collected through cognitive interviews is used to improve questions before sending the survey questionnaire to the full sample (Galindo-Rueda and Van Cruysen, 2016). Cognitive testing is not required for questions and layout features that have previously undergone testing, unless they were tested in a different language or country. Descriptions of the cognitive testing method are provided by Willis (2015, 2005).

9.74.   Respondents do not need to be selected randomly for cognitive testing, but a minimum of two respondents should be drawn from each possible combination of the following three subgroups of the target population: non-innovative and innovative units, service and manufacturing units, and units from two size classes: small/medium (10 to 249 employees) and large (250+ employees). This results in 16 respondents in total. Two (or more) rounds of cognitive testing may be required, with the second round testing revisions to questions made after the first round of testing.

9.75.   In addition to cognitive testing, a pilot survey of a randomly drawn sample from the target population is recommended when substantial changes are made to a questionnaire, such as the addition of a large number of new questions, or new questions combined with a major revision to the layout of the questionnaire. Pilot surveys can help optimise the flow of the questions in the questionnaire and provide useful information on item non-response rates, logical inconsistencies, and the variance of specific variables, which is useful for deciding on sample sizes (see also subsection 9.4.2 below).

9.4. Sampling

9.4.1. The survey frame

9.76.   The units in a survey sample or census are drawn from the frame population. When preparing a survey, the intended target population (for instance all enterprises with ten or more employed persons) and the frame population should be as close as possible. In practice, the frame population can differ from the target population. The frame population (such as a business register) could include units that no longer exist or no longer belong to the target population, and miss units that belong to the target population due to delays in updating the register. The latter can fail to identify small firms with rapid employment growth.

9.77. The frame population should be based on the reference year of the innovation survey. Changes to units during the reference period can affect the frame population, including changes in industrial classifications (ISIC codes), new units created during the period, mergers, splits of units, and units that ceased activities during the reference year.

9.78. NSOs generally draw on an up-to-date official business register, established for statistical purposes, to construct the sample frame. Other organisations interested in conducting innovation surveys may not have access to this business register. The alternative is to use privately maintained business registers, but these are often less up to date than the official business register and can therefore contain errors in the assigned ISIC industry and number of employed persons. The representativeness of private registers can also be reduced if the data depend on firms responding to a questionnaire, or if the register does not collect data for some industries. When an official business register is not used to construct the sampling frame, survey questionnaires should always include questions to verify the size and sector of the responding unit. Units that do not meet the requirements for the sample should be excluded during data editing.

9.4.2. Census versus sample

9.79. While a census will generate more precise data than a sample, it is generally neither possible nor desirable to sample the entire target population, and a well-designed sample is often more efficient than a census for data collection. Samples should always use probability sampling (with known probabilities) to select the units to be surveyed.

9.80. A census may be needed due to legal requirements or when the frame population in a sampling stratum is small. In small countries or in specific sectors, adequate sampling can produce sample sizes for some strata that are close in size to the frame population. In this case using a census for the strata will provide better results at little additional cost. A census can also be used for strata of high policy relevance, such as for large units responsible for a large majority of a country's R&D expenditures or for priority industries. A common approach is to sample SMEs and use a census for large firms.

Stratified sampling

9.81. A simple random sample (one sampling fraction for all sampled units of a target population) is an inefficient method of estimating the value of a variable within a desired confidence level for all strata because a large sample will be necessary to provide sufficient sampling power for strata with only a few units or where variables of interest are less prevalent. It is therefore more efficient to use different sampling fractions for strata that are determined by unit size and economic activity.

9.82. The optimal sample size for stratified sample surveys depends on the desired level of precision in the estimates and the extent to which individual variables will be combined in tabulated results. The sample size should also be adjusted to reflect the expected survey non-response rate, the expected misclassification rate for units, and other deficiencies in the survey frame used for sampling.

9.83. The target sample size can be calculated using a target precision or confidence level and data on the number of units, the size of the units and the variability of the main variables of interest for the stratum. The variance of each variable can be estimated from previous surveys or, for new variables, from the results of a pilot survey. In general, the necessary sample fraction will decrease with the number of units in the population, increase with the size of the units and the variability of the population value, and increase with the expected non-response rate.

9.84.   It is recommended to use higher sampling fractions for heterogeneous strata (high variability in variables of interest) and for smaller ones. The sampling fractions should be 100% in strata with only a few units, for instance when there are only a few large units in an industry or region. The size of the units could also be taken into consideration by using the probability proportional to size (pps) sampling approach, which reduces the sampling fractions in strata with smaller units. Alternatively, the units in each stratum can be sorted by size or turnover and sampled systematically. Different sampling methods can be used for different strata.

9.85.   Stratification of the population should produce strata that are as homogeneous as possible in terms of innovation activities. Given that the innovation activities of units differ substantially by industry and unit size, it is recommended to use principal economic activity and size to construct strata. In addition, stratification by region can be required to meet policy needs. The potential need for age-based strata should also be explored.

9.86.   The recommended size strata by persons employed are as follows:

- small units: 10 to 49

- medium units: 50 to 249

- large units: 250+.

9.87.   Depending on national characteristics, strata for units with less than 10, and 500 or more persons employed can also be constructed, but international comparability requires the ability to accurately replicate the above three size strata.

9.88.   The stratification of units by main economic activity should be based on the most recent ISIC or nationally equivalent industrial classifications. The optimal classification level (section, division, group or class) largely depends on national circumstances that influence the degree of precision required for reporting. For example, an economy specialised in wood production would benefit from a separate stratum for this activity (division 16 of section C, ISIC Rev.4), whereas a country where policy is targeting tourism for growth might create separate strata for division 55 (Accommodation) of section I, for division 56 (Food services) of section I, and for section R (Arts, entertainment and recreation). Sampling strata should not be over-aggregated because this reduces homogeneity within each stratum.

Domains (sub-populations of high interest)

9.89.   Subsets of the target population can be of special interest to data users, or users may need detailed information at industry or regional levels. These subsets are called domains (or sub-populations). To get representative results, each domain must be a subset of the sampling strata. The most frequent approach is to use a high sampling fraction to produce reliable results for the domains. Additionally, establishing domains can allow for the co-ordination of different business surveys, as well as for comparisons over time between units with similar characteristics. Potential sub-populations for consideration include industry groupings, size classes, the region where the unit is located (state, province, local government area, municipality, metropolitan area etc.), R&D-performing units, and enterprise age. Stratification by age can be useful for research on young, innovative enterprises.

9.90.   Relevant preliminary data on domains can be acquired outside of representative surveys run by NSOs, for instance by academics, consultancies or other organisations using surveys or other methods described in the introduction. Academic surveys of start-ups or other domains can produce good results or useful experiments in data collection, as long as they follow good practice research methods.

9.4.3. Longitudinal panel data and cross-sectional surveys

9.91.   As previously noted, innovation surveys are commonly based on repeated cross-sections, where a new random sample is drawn from a given population for each innovation survey. Cross-sectional innovation surveys can be designed in the form of a longitudinal panel that samples a subset of units over two or more iterations of the survey, using a core set of identical questions. Non-core questions can differ over consecutive surveys.

9.92.   Longitudinal panel data allow research on changes in innovation activities at the microeconomic level over time and facilitate research aimed at inferring causal relationships between innovation activities and economic outcomes, such as the share of sales due to innovation (see Chapter 8), by incorporating the time lag between innovation and its outcomes.

9.93.   A number of procedures should be carefully followed when constructing a panel survey:

- Panel units should be integrated within the full-scale cross-sectional survey in order to reduce response burden, maintain an acceptable level of consistency between the results from the two surveys, and collect good quality cross-sectional data for constructing indicators. A panel does not replace the need for a cross-sectional survey.

- Analysis should ensure that the inclusion of results from the panel does not bias or otherwise distort the results of the main cross-sectional survey.

- Panel samples need to be updated on a regular basis to adjust for new entries as well as panel mortality (closure of units, units moving out of the target population) and respondent fatigue. Sample updating should follow the same stratification procedure as the original panel sample.

9.5. Data collection methods

9.94.   Four main methods can be used to conduct surveys: online, postal, computer-assisted telephone interviewing (CATI), and computer-assisted personal interviewing (CAPI or face-to-face interviewing). Online and postal surveys rely on the respondent reading the questionnaire, with a visual interface that is influenced by the questionnaire layout. CATI and face-to-face surveys are aural, with the questions read out to the respondent, although a face-to-face interviewer can provide printed questions to a respondent if needed.

9.95.   The last decade has seen a shift from postal to online surveys in many countries. Most countries that use an online format as their primary survey method also provide a printed questionnaire as an alternative, offered either as a downloadable file (via a link in an e-mail or on the survey site) or via the post.

9.96.   The choice of which survey method to use depends on costs and potential differences in response rates and data quality. Recent experimental research has found few significant differences in either the quality of responses or in response rates, between printed and online surveys (Saunders, 2012). However, this research has mostly focused on households and has rarely evaluated surveys of business managers. Research on different survey methods, particularly in comparison to online formats, is almost entirely based on surveys of university students or participants in commercial web panels. It would therefore be helpful to have more research on the effects of different methods for business surveys.

9.5.1. Postal surveys

9.97.   For postal surveys, a printed questionnaire is mailed to respondents along with a self-addressed postage paid envelope which they can use to return the survey. A best

practice protocol consists of posting a cover letter and a printed copy of the questionnaire to the respondent, followed by two or three mailed follow-up reminders to non-respondents and telephone reminders if needed.

9.98.   Postal surveys make it easy for respondents to quickly view the entire questionnaire to assess its length, question topics, and its relevance. If necessary, a printed questionnaire can be easily shared among more than one respondent, for instance if a separate person from accounting is required to complete the section on innovation expenditures (see section 9.6 below on multiple respondents). A printed questionnaire with filter questions requires that respondents carefully follow instructions on which question to answer next.

9.5.2. Online surveys

9.99.   The best practice protocol for an online survey is to provide an introductory letter by post that explains the purpose of the survey, followed by an e-mail with a clickable link to the survey. Access should require a secure identifier and password and use up-to-date security methods. Follow-up consists of two or three e-mailed or posted reminders to non-respondents, plus telephone reminders if needed.

9.100.   Online questionnaires can be shared, if necessary, among several respondents if the initial respondent provides the username and password to others (see section 9.6).

9.101.   Online surveys have several advantages over postal surveys in terms of data quality and costs:

- The software can notify respondents through a pop-up box if a question is not completed or contains an error, for instance if a value exceeds the expected maximum or if percentages exceed 100%. With a postal survey, respondents need to be contacted by telephone to correct errors, which may not occur until several weeks after the respondent completed the questionnaire. Due to the cost of follow-up, missing values in a postal survey are often corrected post-survey through imputation.

- Pop-up text boxes, placed adjacent to the relevant question, can be used to add additional information, although respondents rarely use this feature.

- Respondents cannot see all questions in an online survey and consequently are less likely than respondents to a printed questionnaire to use a "no" response option to avoid answering follow-on questions. An online survey can therefore reduce false negatives.

- Survey costs are reduced compared to other survey methods because there is less need to contact respondents to correct some types of errors, data are automatically entered into a data file, data editing requirements are lower than in other methods, and there are reduced mailing and printing costs.

9.102.   The main disadvantage of an online survey compared to other survey methods is that some respondents may be unable to or refuse to complete an online form. In this case an alternative survey method is required (see subsection 9.5.4 below). The online system may also need to be designed so that different persons within a unit can reply to different sections of the survey.

Collecting paradata in online surveys

9.103.   Online survey software offers the ability to collect paradata on keystrokes and mouse clicks (for instance to determine if help menus have been accessed) and response

time data, such as the time required to respond to specific questions, sections, or to the entire survey (Olson and Parkhurst, 2013). Paradata can be analysed to identify best practices that minimise undesirable respondent behaviour such as premature termination or satisficing, questions that are difficult for respondents to understand (for instance if question response times are considerably longer than the average for a question of similar type), and if late respondents are more likely than early ones to speed through a questionnaire, thereby reducing data quality (Belfo and Sousa, 2011; Fan and Yan, 2010; Revilla and Ochoa, 2015).

9.104. It is recommended to **collect paradata when using online surveys** in order to identify issues with question design and questionnaire layout.

9.5.3. Telephone and face-to-face interviews

9.105. Telephone and face-to-face surveys use computer-assisted data capture systems. Both methods require questions to be heard, which can require changes to question formats compared to visual survey methods. Interviewers must be trained in interview techniques and how to answer questions from the respondent, so that the respondent's answers are not biased through interactions with the interviewer. For both formats, filters are automatic and the respondent does not hear skipped questions, although interviewers can probe for additional information to ensure that a "no" or "yes" response is accurate.

9.106. The CATI method has a speed advantage compared to other methods, with results obtainable in a few weeks. Both CATI and CAPI can reduce errors and missing values, as with online surveys. Their main disadvantage, compared to an online survey, is higher costs due to the need for trained interviewers. Secondly, compared to both online and postal surveys, CATI and CAPI methods are unsuited for collecting quantitative data that require the respondent to search records for the answer.

9.107. The main reason for using the CAPI format is to obtain high response rates. This can occur in cultures where face-to-face interviews are necessary to show respect to the respondent and in areas where online or postal surveys are unreliable.

9.5.4. Combined survey methods

9.108. The use of more than one survey method can significantly increase response rates (Millar and Dillman, 2011). Where possible, surveys should combine complementary survey methods that are either visual (printed or on line) or aural (CATI or face-to-face) because of differences by survey methods in how respondents reply to questions. Telephone surveys can also elicit higher scores than online or mailed surveys on socially desirable questions (Zhang et al., 2017). As innovation is considered socially desirable, this could result in higher rates of innovation reported to CATI surveys compared to printed or online surveys. Possible survey method effects should be assessed when compiling indicators and comparing results across countries that use different survey methods.

9.6. Survey protocol

9.109. The survey protocol consists of all activities to implement the questionnaire, including contacting respondents, obtaining completed questionnaires, and following up with non-respondents. The protocol should be decided in advance and designed to ensure that all respondents have an equal chance of replying to the questionnaire, since the goal is to maximise the response rate. Nonetheless, the optimum survey protocol is likely to vary by country.

9.6.1. Respondent identification

9.110. Choosing a suitable respondent (or department within a large firm) is particularly important in innovation surveys because the questions are specialised and can be answered by only a few people, who are rarely the same as the person who completes other statistical questionnaires. In small units, managing directors are often good respondents. As much as possible, knowledgeable respondents should be selected to minimise the physical or virtual "travel" of a questionnaire to multiple people within a firm. Travel increases the probability that the questionnaire is lost, misplaced or that no one takes responsibility for its completion. In large units where no single individual is likely to be able to respond to all questions, some travel will be inevitable. However, a single, identified contact person or department should be responsible for co-ordinating questionnaire completion.

9.6.2. Support for respondents

9.111. Innovation surveys contain terms and questions that some respondents may not fully understand. Survey managers need to train personnel to answer potential questions and provide them with a list of basic definitions and explanations of questions.

9.6.3. Mandatory and voluntary surveys

9.112. Completion of innovation surveys can be either voluntary or mandatory, with varying degrees of enforcement. Higher non-response rates are expected for voluntary surveys and are likely to increase with questionnaire length. Sampling fractions can be increased to account for expected non-response rates, but this will not solve potential bias due to differences in the characteristics of non-respondent and respondent units that are correlated with survey questions. Reducing bias requires maximising response rates and representativeness (see below).

9.113. Whether a survey is voluntary or mandatory can also affect results. For example, the calculated share of innovative firms in a voluntary survey will be biased upwards if managers from non-innovative firms are less likely to respond than the managers of innovative firms (Wilhelmsen, 2012).

9.6.4. Non-response

9.114. Unit non-response occurs when a sampled unit does not reply at all. This can occur if the surveying institute cannot reach the reporting unit or if the reporting unit refuses to answer. Item non-response refers to the response rate to a specific question and is equal to the percentage of missing answers among the responding units. Item non-response rates are frequently higher for quantitative questions than for questions using nominal or ordinal response categories.

9.115. Unit and item non-response are only minor issues if missing responses are randomly distributed over all units sampled and over all questions. When unit non-responses are random, statistical power can be maintained by increasing the sampling fraction. When item non-responses are random, simple weighting methods can be used to estimate the population value of a variable. However, both types of non-response can be subject to bias. For example, managers from non-innovative units could be less likely to reply because they find the questionnaire of little relevance, resulting in an overestimate of the share of innovative units in the population. Or, managers of innovative units could be less likely to reply due to time constraints.

Improving response rates

9.116. Achieving high response rates, particularly in voluntary surveys, can be supported by good question and questionnaire design (see section 9.3) as well as good survey protocols. Two aspects of the survey protocol can have a large positive effect on response rates: (i) good follow-up with multiple reminders to non-respondents; and (ii) the personalisation of all contacts, such as using the respondent's name and changing the wording of reminder e-mails. Personalisation includes sending a first contact letter by post, which can significantly increase response rates in comparison to a first contact via e-mail (Dykema et al., 2013). Clearly communicating the purpose and use of the survey data is critical for generating trust and participation. Participation can be further enhanced if managers anticipate direct benefits for their business from providing truthful and carefully thought-out answers.

Managing low unit response rates

9.117. There are no clear boundaries for high, moderate and low unit response rates. The rule of thumb is that high response rates exceed 70% or 80%, moderate response rates are between 50% and 70% or 80%, and low response rates are below 50%.

9.118. Unless the response rate is very high (above 95%), differences between respondents and non-respondents should be compared using stratification variables such as unit size or industry. If the response rate is high and there are no significant differences by stratification variables, population weighting can be calculated on the basis of the units that replied. This procedure assumes that the innovation behaviour of responding and non-responding units conditional on these characteristics is identical. Challenges can arise when behaviour is very heterogeneous within strata (e.g. between large and very large firms).

9.119. If the response rate is moderate or low, it is recommended to conduct a non-response survey (see subsection 9.6.5 below).

9.120. If the unit response rate is very low (below 20%), a non-response survey may be insufficient for correcting for potential bias, unless it is of very high quality and covers a large share of non-responding units. The data can be analysed to determine if response rates are acceptable in some strata and to conduct a non-response survey for those strata. Otherwise, the results should not be used to estimate the characteristics of the target population because of the high possibility of biased results. It is possible to use the data to investigate patterns in how variables are correlated, as long as the results are not generalised to the target population.

9.6.5. Conducting non-response surveys

9.121. Many NSOs have their own regulations for when a non-response survey is necessary. Otherwise, a non-response survey is recommended when the unit non-response rate in a stratum exceeds 30%. The non-response survey should sample a minimum of 10% of non-respondents (more for small surveys or for strata with a low population count).

9.122. The objective of the non-response survey is to identify significant differences between responding and non-responding units in innovation activities. To improve future surveys, it is possible to obtain information on why non-respondent units did not answer. In the ideal case, the unit response rate for the non-response survey is sufficiently high and replies are sufficiently reliable to be useful for adjusting population weightings. However, survey method effects in the non-response survey (different survey methods or questionnaires compared to the main survey) should also be considered when adjusting population weights.

9.123. The non-response survey questionnaire must be short (no more than one printed page) and take no more than two to three minutes to complete. The key questions should replicate, word for word, "yes or no" questions in the main survey on innovation outputs (product innovations and business process innovations) and for some of the innovation activities (for instance R&D, engineering, design and other creative work activities, etc.). If not available from other sources, the non-response survey needs to include questions on the unit's economic activity and size.

9.124. Non-response surveys are usually conducted by CATI, which provides the advantage of speed and can obtain high response rates for a short questionnaire, as long as all firms in the sample have a working contact telephone number. The disadvantage of a CATI survey as a follow-up to a postal or online survey is that short telephone surveys in some countries could be more likely than the original survey to elicit positive responses for questions on innovation activities and outputs. The experience in this regard has been mixed, with different countries obtaining different results. More experimental research on the comparability of business survey methods is recommended.

9.7. Post-survey data processing

9.125. Data processing involves checks for errors, imputation of missing values and the calculation of weighting coefficients.

9.7.1. Error checks

9.126. As noted in subsections 9.5.2 and 9.5.3 above, the use of online, CATI and CAPI survey methods can automatically identify potential errors and request a correction from respondents. All of the following types of error checks are required for printed questionnaires, but only the check for out of scope units might be required for an online survey. When errors are identified, the respondent or reporting unit should be contacted as soon as possible to request a correction.

Out of scope units

9.127. Responses can be obtained from out of scope units that do not belong to the target population, such as a unit that is below the minimum number of persons employed, a unit that is not owned by a business, or a unit in an excluded ISIC category. Responses from these units must be excluded from further analysis.

Data validation checks

9.128. These procedures test whether answers are permissible. For example, a permissible value for a percentage is between 0 and 100.

9.129. An additional check for data quality should be applied to ratio and interval level data, particularly for innovation expenditures. Current best practice is to compare interval level data against other available sources (for instance for R&D expenditures and for expenditures on capital equipment). In addition, estimates of the innovation sales share and other interval level data should be checked for outliers or other unexpected values. These methods are particularly important for large units that account for a high share of total reported R&D and innovation expenditures.

Relational checks

9.130. These evaluate the relationship between two variables and can identify hard and soft errors. Hard errors occur when a relationship must be wrong, for instance if percentages do not sum to 100% or if the number of reported persons employed with a tertiary education exceeds the total reported number of persons employed. Other relational checks identify soft errors where a response could be wrong. For instance, a unit with ten persons employed could report EUR 10 million of innovation expenditures. This is possible, but unlikely.

Routing error checks

9.131. These checks test whether all questions that should have been answered have been answered, i.e. respondents from innovation-active units answered all questions on innovation expenditures. An error indicates that the respondent did not understand or follow the filtering instructions.

9.7.2. *Imputation of missing data*

9.132. Another type of error is when a respondent fails to answer a question, for instance several sub-questions in a grid question are left blank, either intentionally or accidentally. Respondents can also refuse to answer a question if they find that none of the response categories are appropriate or if they don't know the answer and a "don't know" option is not provided. Online survey software can force the respondent to reply, but this is not recommended for voluntary surveys because it may cause the respondent to abandon the questionnaire.

9.133. To reduce costs and response burden, missing values can in some cases be imputed through the use of additional information, instead of re-contacting the respondent. The use of additional information should provide a more accurate estimate of missing values than simply using the mean observed value in a stratum. Imputed values should always be flagged to prevent their possible use as dependent values in multivariate analyses to avoid biased coefficients.

9.134. Cold-deck imputation techniques estimate missing values using data in other statistical surveys (including previous surveys) or from other related sources. For instance, data on the number of employees with a tertiary education could be available in a separate survey for similar types of units.

9.135. Hot-deck imputation uses other data from the innovation survey to impute some missing values. The choice of which hot-deck method to use depends on the measurement level for the variable. Interval level data can be imputed either by using the mean value of the variable in the responding unit's stratum, or by using regression to predict the value of the interval level variable. In the latter case, the results need to be checked to identify non-credible estimated values, such as negative values.

9.136. Nearest-neighbour techniques can be used to impute missing nominal and ordinal level values. This technique uses data from clean records (a donor case with a record not violating any error check), in order to replace the missing value with the value in the donor record. The donors are chosen on the basis of similar stratification variables and for maximum comparability on related variables. For example, the donor for a missing ordinal variable on collaboration partners should be as similar as possible for related variables on information sources.

9.7.3. Calculating weights

9.137. The results of sample surveys need to be weighted to provide estimates that are representative of the target population. There are various methods for weighting sampling results. The simplest is weighting by the reciprocal (inverse) of the sampling fractions of the stratum for each sampled unit, corrected by the unit non-response for each stratum, which gives the realised sample. For instance, if the sampling fraction is 10/100, but 10% of the sampled units did not respond, the corrected sampling fraction is 9/100. This procedure assumes that the distribution of innovation activities among responding and non-responding units is identical. This assumption can be tested through a non-response analysis, and even if the assumption is wrong, the bias introduced can be disregarded if the fraction of non-responding units is fairly small.

9.138. Not all sample surveys use stratification – a census by definition is not stratified. For a non-stratified survey, weighting should use the reciprocal of the overall sampling fraction, adjusted for the reciprocal of the total response rate. For a census, strata can be constructed post-survey to identify non-response rates in strata defined by firm size, sector, region etc. Census weighting variables can be constructed from the reciprocal of the strata response rates.

9.139. The final weighting factors should be further corrected if a non-response survey identifies statistically significant differences between the original survey respondents and the respondents to the non-response survey, for instance if a lower percentage of non-innovative units replied to the full survey compared to the non-respondent survey. One approach is to divide each stratum into a number of response homogeneity groups (RHGs) with (assumed) equal response probabilities within groups. RHGs can be determined using the results of the non-response survey. A second approach is to use auxiliary information at the estimation stage for reducing the non-response bias or two-phase sampling estimation methods. In the latter case, the sample is split by the phenomenon for which a likely non-response bias has been investigated (e.g. innovative versus non-innovative firms) and weighting factors are calculated separately for each group. In a second step, weighting factors are adjusted by a non-response correction factor that represents the bias between responding and non-responding firms with respect to the investigated phenomenon.

9.140. The weights can be further refined by calibration if the frame population includes some quantitative or qualitative information on all units, such as number of employees, turnover, legal status or region. The calibration ensures that the weighted sample sums to the total population or distribution for the frame variables and can increase precision and reduce bias. Effective calibration software, in particular CLAN from Statistics Sweden, CALMAR from the French National Statistical Insitute (INSEE) in France, and G-Est from Statistics Canada, are available for use by other countries. Many of the software packages used for calculating weights can also calculate measures of sampling variability.

9.8. Publication and dissemination of results

9.141. Innovation surveys are used to produce tables of innovation statistics and indicators and in econometric analyses of a variety of topics concerning innovation. The production of statistics and indicators requires using population weights to produce representative results for the target population. Most innovation surveys use a probability sample for many strata. Surveys can create two types of errors for indicators: random errors due to the random process used to select the units, and systematic errors containing all non-random errors (bias). The probability of random errors should be provided with the results by

including the confidence intervals, standard errors and coefficients of variation where applicable. Confidence limits span the true but unknown values in the survey population with a given probability. If possible, data quality reports should also provide an evaluation of non-random errors.

9.8.1. Metadata and quality reports

9.142. The presentation of statistics and indicators should contain metadata, including information on the procedure used to collect data, sampling methods, procedures for dealing with non-response, and quality indicators. This will allow users to better interpret the data and judge its quality. International organisations should continue to provide detailed information about the common and idiosyncratic methodologies adopted by countries covered in their databases and reports.

9.8.2. Data access

9.143. Descriptive data can be provided through press releases, tables, databases and reports. Econometric analysis of innovation survey data is of considerable value to policy development (see section 11.5), but this is not the main task of NSOs. In-house econometric analyses by NSOs can be supplemented at low cost by providing researchers with access to innovation survey microdata. This requires maintaining confidentiality, either through the provision of a safe centre for data access by external academics or through the construction of anonymised data sets.

References

Barge, S. and H. Gehlbach (2012), "Using the theory of satisficing to evaluate the quality of survey data", *Research in Higher Education*, Vol. 53/2, pp. 182-200.

Belfo, F.P. and R.D. Sousa (2011), "A web survey implementation framework: evidence-based design practices", conference paper for the 6th Mediterranean Conference on Information Systems, MCIS 2011, Limassol, 3-5 September, http://aisel.aisnet.org/mcis2011/43/.

Cirera, X. and S. Muzi (2016), "Measuring firm-level innovation using short questionnaires: Evidence from an experiment", *Policy Research Working Papers*, No. 7696, World Bank Group.

Couper, M.P. et al. (2013), "The design of grids in web surveys", *Social Science Computer Review*, Vol. 31/3, pp. 322-345.

Downes-Le Guin, T. et al. (2012), "Myths and realities of respondent engagement in online surveys", *International Journal of Market Research*, Vol. 54/5, pp. 613-633.

Dykema, J. et al. (2013), "Effects of e-mailed versus mailed invitations and incentives on response rates, data quality, and costs in a web survey of university faculty", *Social Science Computer Review*, Vol. 31/3, pp. 359-370.

EC et al. (2009), *System of National Accounts 2008*, United Nations, New York, https://unstats.un.org/unsd/nationalaccount/docs/sna2008.pdf.

Fan, W. and Z. Yan (2010), "Factors affecting response rates of a web survey: A systematic review", *Computers in Human Behavior*, Vol. 26/2, pp. 132-139.

Galesic, M. and M. Bosnjak (2009), "Effects of questionnaire length on participation and indicators of response quality in a web survey", *Public Opinion Quarterly*, Vol. 73/2, pp. 349-360.

Galindo-Rueda, F. and A. Van Cruysen (2016), "Testing innovation survey concepts, definitions and questions: Findings from cognitive interviews with business managers", *OECD Science, Technology and Innovation Technical Papers*, OECD Publishing, Paris, http://oe.cd/innocognitive.

Harkness, J.A. et al. (eds.) (2010), *Survey Methods in Multicultural, Multinational, and Multiregional Contexts*, Wiley Series in Survey Methodology, John Wiley & Sons, Hoboken.

Harris, R.I.D. (1988), "Technological change and regional development in the UK: Evidence from the SPRU database on innovations", *Regional Studies*, Vol. 22/5, pp. 361-374.

Hoskens, M. et al. (2016), "State of the art insights in capturing, measuring and reporting firm-level innovation indicators", paper for the OECD Blue Sky 2016 Forum, www.oecd.org/sti/069%20-%20Measuring%20innovation_ECOOM%20August%202016.pdf.

Kleinknecht, A., J.O.N. Reijnen and W. Smits (1993), "Collecting literature-based innovation output indicators: The experience in the Netherlands", in *New Concepts in Innovation Output Measurement*, Palgrave Macmillan, London, pp. 42-84.

Millar, M.M. and D.A. Dillman (2011), "Improving response to web and mixed-mode surveys", *Public Opinion Quarterly*, Vol. 75/2, pp. 249-269, https://doi.org/10.1093/poq/nfr003.

OECD (2015a), *Recommendation of the OECD Council on Good Statistical Practice*, OECD, Paris, www.oecd.org/statistics/good-practice-toolkit/Brochure-Good-Stat-Practices.pdf.

OECD (2015b), *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development*, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris, http://oe.cd/frascati.

Olson, K, and B. Parkhurst (2013), "Collecting paradata for measurement error evaluations", in *Improving Surveys with Paradata: Analytic Uses of Process Information*, John Wiley & Sons, Hoboken, pp. 43-72.

Revilla, M. and C. Ochoa (2015), "What are the links in a web survey among response time, quality and auto-evaluation of the efforts done?", *Social Science Computer Review*, Vol. 33/1, pp. 97-114, https://doi.org/10.1177/0894439314531214.

Saunders, M.N.K. (2012), "Web versus mail: The influence of survey distribution mode on employees' response", *Field Methods*, Vol. 24/1, pp. 56-73.

Snijkers, G. and D.K. Willimack (2011), "The missing link: From concepts to questions in economic surveys", paper presented at the 2nd European Establishment Statistics Workshop (EESW11), Neuchâtel, Switzerland, Sept. 12-14.

Snijkers, G. et al. (eds.) (2013), *Designing and Conducting Business Surveys*, Wiley Series in Survey Methodology, John Wiley & Sons, Hoboken.

Tourangeau, R., L.J. Rips and K. Rasinski (2000), *The Psychology of Survey Response*, Cambridge University Press, Cambridge.

UN (2008)，*International Standard Industrial Classification of All Economic Activities (ISIC)，Revision 4*，United Nations, New York, https://unstats.un.org/unsd/publications/catalogue?selectID=396.

UN (2007), *Statistical Units*, United Nations, New York, http://unstats.un.org/unsd/isdts/docs/StatisticalUnits.pdf.

Wilhelmsen, L. (2012), "A question of context: Assessing the impact of a separate innovation survey and of response rate on the measurement of innovation activity in Norway", *Documents*, No. 51/2012, Statistics Norway, Oslo, www.ssb.no/a/english/publikasjoner/pdf/doc_201251_en/doc_201251_en.pdf.

Willeboordse, A. (ed.) (1997), *Handbook on Design and Implementation of Business Surveys*, Eurostat, Luxembourg, http://ec.europa.eu/eurostat/ramon/statmanuals/files/Handbook%20on%20surveys.pdf.

Willis, G.B. (2015), *Analysis of the Cognitive Interview in Questionnaire Design*, Oxford University Press, Oxford.

Willis, G.B. (2005), *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, SAGE Publications.

Zhang, X.C. et al. (2017), "Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ", *Computers in Human Behavior*, Vol. 71, pp. 172-180.

Oslo Manual 2018
Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition