

Handbook on Registers-Based Population and Housing Censuses

DRAFT

(Version of 22 November 2021)

Please do not circulate or cite without permission



United Nations
Statistics Division

Table of Contents

1. Introduction.....	5
1.1. Background, including the rationale for moving to a register-based statistical system.....	5
1.2. Overview of the handbook and target audience.....	7
1.3. What is a register?	10
1.4. Definition and essential features of the census	12
1.4.1 <i>Definition of the population census</i>	12
1.4.2 <i>Definition of the housing census</i>	12
1.4.3 <i>Essential features of the census</i>	13
1.5 Data collection methodologies.....	15
1.5.1 <i>The traditional census</i>	16
1.5.2 <i>Register-based census</i>	21
1.5.3 <i>Combined approach</i>	24
1.6 Current profile of global censuses	29
2 Moving to a register-based approach.....	35
2.1 Introduction.....	35
2.2 Necessary pre-conditions.....	36
2.2.1 <i>NSO's long-term project vision for a register-based statistical system</i>	36
2.2.2 <i>Legal framework and role of the NSO to be established</i>	37
2.2.3 <i>Political support</i>	38
2.2.4 <i>Public approval</i>	39
2.2.5 <i>Cooperation between NSO and register-holders</i>	40
2.2.6 <i>Organizational frameworks</i>	41
2.2.7 <i>Knowledge of administrative registers and initial assessment of their</i> <i>quality</i>	42
2.2.8 <i>Unified identification system</i>	43
2.2.9 <i>Institutional infrastructure and capacity</i>	44
2.3 Census-specific pre-conditions	45
2.3.1 <i>Census-specific legislation</i>	45
2.3.2 <i>Data user requirements</i>	46
2.3.3 <i>Confidentiality and public approval</i>	47
2.3.4 <i>Knowledge of relevant administrative sources</i>	48
2.3.5 <i>Capacity for assessing the availability and quality of administrative registers</i>	49
2.3.6 <i>Availability of base registers for censuses</i>	49
2.3.7 <i>Identifiers</i>	52
2.4 Difficulties and challenges – and solutions	52
2.4.1 <i>Greater dependency on public authorities</i>	52
2.4.2 <i>Differences in concepts and definitions</i>	54

2.4.3 Classification systems.....	55
2.4.4 Inconsistency between sources.....	57
2.4.5 Timeliness and reference periods	58
2.4.6 Lack of a unified identification system	60
2.4.7 Missing data and the under-coverage of some populations	60
2.4.8 Maintaining IT infrastructure	61
2.4.9 Privacy and security concerns	62
2.4.10 Resistance to change.....	63
2.4.11 Diminishing interest	64
2.5 Impact on the essential features of the census	64
2.5.1 Individual enumeration.....	64
2.5.2 Simultaneity	65
2.5.3 Universality.....	66
2.5.4 Defined periodicity	66
2.5.5 Small area data	67
3 Types of administrative sources	69
3.1 Introduction.....	69
3.2 The base registers	70
3.2.1 Population registers	70
3.2.2 Building and dwelling registers	85
3.2.3 Enterprise (business) registers	86
3.2.4 Address registers	88
3.3 Specialised or supplementary registers	89
3.3.1 Social security or pension registers	90
3.3.2 Tax registers	90
3.3.3 Employment, unemployment and jobseeker registers	90
3.3.4 Education and student registers	91
3.3.5 Health registers	91
3.3.6 Border control data	91
3.3.7 Other supplementary registers	92
3.4 Options if data are not available from administrative sources	93
3.5 Big data.....	96
4.1 Introduction.....	99
4.2 Creating a statistical population register	100
4.3 Constructing integrated statistical registers for the purpose of	102
census	102
4.3.1 Data linkage	103
4.3.2 Dealing with duplication in the statistical population register	109

4.3.3 Conflict resolution	109
4.3.4 Updating and the 'Signs of Life' methodology.....	110
4.3.5 Editing and imputation	114
4.4 Research and testing	117
5 Quality assessment	119
5.1 Introduction.....	119
5.2 Quality dimensions.....	120
5.2.1 Relevance.....	120
5.2.2 Accuracy and reliability.....	122
5.2.3 Timeliness and punctuality.....	123
5.2.4 Coherence and comparability.....	123
5.2.5 Accessibility and interpretability	124
5.3 Stages of quality assessment	125
5.3.1 Source quality.....	126
5.3.2 Input quality	139
5.3.3 Process quality.....	143
5.3.4 Output quality	150
5.4 Reporting quality.....	155
6 Conclusions and recommendations.....	156
6.1 Summary conclusions	156
6.2 Recommendations	158
Annex A Definitions of terms and concepts.....	161
Annex B References	171

1. Introduction

1.1. Background, including the rationale for moving to a register-based statistical system

1-1 Population and housing censuses play a vital role in any national statistical system in the production of official statistics on a wide range of topics related to the demographic, social and economic characteristics of the population and its housing. There are various approaches to collecting the data for the purposes of the census, involving either the traditional collection directly from the public in a field enumeration, or the use of existing administrative registered-based data sources – or a combination of both. But, regardless of methodological approaches, the crucial principle of providing detailed, multi-variate statistics, particularly at lowest geographic level, remains of paramount importance.

1-2 Administrative registers generated by governments and service providers are becoming an increasingly important source of data for official statistics, and particularly for producing census information. Using administrative sources offers many opportunities in conducting censuses, in particular for reducing costs, improving the quality of data and producing more frequent information—in many cases on a yearly basis. Such potential may be particularly relevant for the purposes of the annual monitoring of Sustainable Development Goal (SDG) indicators. But there are also many challenges relating, for example, to the quality of existing administrative registers, the lack of a necessary legal framework to authorise the use of individual records for statistical production, and the institutional capacity for establishing and maintaining statistical registers.

1-3 Implementation of the 2020 round of population and housing censuses has been significantly disrupted by the arrival of the Covid-19 pandemic. National Statistical Offices (NSOs) have been challenged in their attempt to conduct their census as scheduled in 2020 and 2021 to the extent that many countries that use traditional census or combined censuses with a large scale of field data collection have had to postpone their censuses for more than a year. As a consequence, in order to reduce the impact of the Covid-19 pandemic on the implementation of 2020 round of censuses (with a view, in particular, to

avoid direct contact with the public), there is now more interest in the use of administrative data and online data collection methods.

1-4 Changing the design of the census processes – particularly the data collection operation - from traditional methods to other approaches that rely on the use of administrative data is not easy task and can be more complex compared to the well-established design of a traditional census. When census information is produced using only administrative registers (as has been the case in the Nordic countries for several decades), the approach is almost unique to those countries that basically collect information on individuals, households and dwellings from existing administrative sources through linking different types of registers, of which population registers, building and dwelling registers, business registers, tax registers, education registers, unemployment registers and social security registers are of primary importance².

1-5 When administrative registers are used in combination with other data collection methods, census processes and procedures can vary significantly from one country to another. Depending on the quality and availability of administrative sources and expertise in the use and management of administrative data, the design of the census processes and procedures can be sometimes unique to an individual country. For example, countries which do not have reliable population registers may use the combined census with full-field enumeration for different purposes, including:

- (i) improving the accuracy of population count,
- (ii) validating, checking and updating the data derived from administrative sources,
- (iii) linking different administrative registers, and/or
- (iv) adding new variables to census database.

1-6 Alternatively, when administrative registers are combined with an ad-hoc survey or with existing household surveys, the design of census procedures will be completely different compared to the combined method with full-field enumeration. Therefore, it is a challenge to categorize these approaches using generally accepted data collection methods and census methodologies.

1-7 It should be noted that current international guidelines^{1,2} provide only basic information for adopting different approaches in conducting censuses with

the use of registers, and need to be developed in order to provide better guidance on possible options for designing census processes and procedures that use of administrative data. The aim of this handbook is, therefore, to provide a more detailed discussion on the census methodologies that rely on the use of administrative registers either as the sole source of input data or in a combined approach. The handbook aims also to elaborate on the key considerations and requirements for transitioning from a traditional census to other census methodologies and the importance of building institutional capacity and infrastructure for maintaining statistical registers for the successful implementation of such a new methodology. At the more basic level, the handbook may be particularly helpful as a reference to countries starting from scratch or experimenting on the use of administrative registers for the purpose of censuses.

1-8 In doing so the handbook builds on the work and outputs that have been prepared by the Conference of European Statistics, most notably *Guidelines on the use of registers and administrative data for population and housing censuses* and *Guidelines for assessing the quality of administrative sources in censuses*; it also incorporates practices, challenges and lessons learned by those NSO's with previous experience in the use of administrative data for producing census information.

1.2. Overview of the handbook and target audience

1-9 Building on the experiences of those countries that have successfully done so, the handbook is targeted at those NSOs that similarly wish to transition from their traditional census methodology to one where the use of data derived from administrative registers can be successfully developed and implemented for future censuses as well as being integrated with their statistical systems more generally. A number of key components to such a transition will be discussed. These include:

Necessary pre-conditions for the use of administrative registers for population and housing censuses

1-10 The conditions that should be met before NSOs should attempt to use information from administrative registers will usually include: having a clear vision and long-term strategy for a register-based statistical system; government

support; an appropriate legal framework in which the role of the NSO is clearly established; public and stakeholder approval; cooperation between the NSO and the relevant register-holders; knowledge of administrative registers and quality assessment of base registers; a unified identification system; and an appropriate institutional infrastructure and capacity. These are discussed in detail in Chapter 2 - **Moving to a register-based approach**, together with an exposition of the advantages of moving to a register-based statistical system and the difficulties and challenges in doing so.

Establishing base and supplementary registers

1-11 Chapter 3 - **Types of administrative sources** discusses the role and importance of:

(a) the base registers for used in censuses - the registers of population, buildings and dwellings, and of business establishments and enterprises; and

(b) supplementary registers (sometimes referred to as 'specialized registers) such as education, social security or unemployment registers, in providing specific topic-oriented data not held in the base registers.

1-12 The options if particular data are not available from administrative sources are also discussed. For example, when information for some census topics is not held in any administrative register or the quality is not sufficiently good, information can be derived by integrating data collected from existing, or ad-hoc, surveys.

Building comprehensive and reliable statistical registers for censuses and combining with surveys

1-13 Depending on the quality and content of these registers, different approaches for transitioning to a register-based census methodology are discussed in Chapter 4 – **The transition process**. For example, a full-field enumeration can be used to improve the quality of population register and establish a link between persons and the address of their usual residence, or such a link may otherwise be established based on previous censuses, civil registration and register of unique personal identification number or through linking existing reliable administrative registers.

1-14 Chapter 4 also discusses the linking of base registers with the supplementary or specialized registers to produce a statistical database from which output tabulations can be prepared. Such linkage offers a great opportunity for producing statistics on a wide range of census topics as well as for improving the coverage and content of the population register. When register data are used for statistical purposes, and especially when data from several registers are combined, problems concerning consistency and data quality may be revealed.

Assessment of the quality of administrative data sources

1-15 An assessment of the quality of the census data has always been an important and necessary task. Assessing the quality of a census that makes use of a new data collection methodology is especially important, as it provides relevant information on the reliability of the new census results, and how the quality may differ from the results of previous censuses.

1-16 Chapter 5 – **Quality assessment** discusses the process and tools by which the quality of data sources and the individual data items provided by data holders can be assessed. For example, a dual system estimation approach might be used to measure coverage errors in the population register. Another approach is to compare population register with a set of reliable registers or other administrative data sets for checking the coverage errors of a census (through identifying ‘signs of life’ activities). These approaches would be supported by other quality assurance processes, for example demographic analysis and comparisons with household surveys or other data sources. The impact on the quality of the data as a result of the data processing operation within the NSO is also covered.

Conclusions

1-17 It should be stressed that the handbook is not intended to set out specific recommendations on how NSOs should (or should not) approach moving to an administrative data-based census. Instead, Chapter 4 summarises some of the processes and actions that NSO may wish to adopt in order to maximise the likelihood of a successful transition to such a census and how to resolve the difficulties and challenges in doing so, based on the experience of others.

Case Studies (in progress)

1-18 A number of case studies that will be prepared by the NSOs themselves will be included in Appendix to illustrate how particular issues and challenges were met by some of those countries that have already transitioned to a register-based or combined census, or are in the process of preparing to do so to.

1-19 The rest of this introductory Chapter sets the scene for the handbook by attempting to define the concept of a register, setting out the essential features of the census, and discussing the various methodologies for collecting census data currently being adopted by different countries. A glossary of definitions of terms and concepts used throughout the handbook is also included as an annex.

1.3. What is a register?

1-20 In the most general sense, a *register* is a systematic collection of unit-level data organized in such a way that updating is possible, where ‘updating’ is the process of identifiable information with the purpose of establishing, bringing up to date, correcting or extending the register - that is, keeping track of any changes in the data describing the units and their attributes.

1-21 Administrative registers are produced for the purpose of effecting administrative processes by recording information on units (such as persons and dwellings) and variables that are defined by administrative rules and demands in a country. Although the content and process of registers will differ from one country to another, the types of the registers maintained are usually very similar. The following definitions of the main concepts used in the system of administrative registers refer to the UNECE’s 2007 report on register-based statistics as developed in the Nordic countries⁵. The more detailed descriptions of the various registers that are used for the purposes of the census are set out in Chapter 3.

1-22 As a rule, a register will contain information on a complete group of units, referred to as the *target population* (such as persons, dwellings, buildings or business organisations). Such units are defined by a precise set of rules in order to fulfil the function of the register (for example, in order to identify the *resident population in a country*), and the attributes are updated in line with changes affecting the units.

1-23 Information in registers can be stored in a variety of ways. It should be stressed, however, that for use in censuses (and indeed, any for statistical

purpose) in the data should be computer-based and available to the NSO in a machine-readable (digital) form.

1-24 A key requirement is that each unit in the register should always be uniquely identified. This is best achieved by using a system of *identification codes* (keys), but identification is also possible without such a code if sufficient information on the units is available (for example, for persons such information would relate to name, sex, address and date of birth). The role of such unique identifiers will be discussed in Chapter 4.

1-25 *Administrative registers* are registers primarily used in an administrative information system by public agencies (such as government or local authority bodies) or private institutions or companies for their own specific purposes (such as the monitoring of taxes, the allocation of pensions, the provision of services, or the production of goods). In this context, the term '*administrative data*' refers to a set of units and data derived from an administrative source. It follows from this definition that an '*administrative data source*' is a data holding that contains information collected primarily for administrative, rather than for research or statistical, purposes.

1-26 *Administrative base registers* are kept as a basic resource for public administration. The function is to record the stock of the population and to maintain identification information. *Statistical base registers* are based on the corresponding administrative registers. Their principal tasks are to define important populations and contain links to other base registers.

1-27 *Specialized (or supplementary) registers* are registers that, unlike base registers, serve one or more specific purposes (such as the administration of health or educational services or the allocation of state benefits). They often hold information on the population and some basic data from a base register but contain other data that are particular to their specific administrative purpose and which not held in other registers (such as a register of licensed vehicle owners).

1-28 *Statistical registers* are created (normally by statisticians) by processing data from administrative registers for statistical purposes in accordance with statistical concepts and definitions. A statistical register could be based on one or more administrative registers. Statistical registers are also referred to as *secondary registers*.

1-29 The distinction between administrative and statistical registers should always be borne in mind, and particularly when the topic of population registers is discussed in Chapter 3. A full definition of a population register is included in that Chapter but, for the time being, it should be sufficient just to note here that a '*population register*' is a register that provides a frame of residents in a given territory that also provides some basic demographic characteristics of individual persons.

1.4. Definition and essential features of the census

1-30 The definition and essential features of the census are clearly set out in the 3rd revision of the UN's *Principles and Recommendations for Population and Housing Census*¹, but it is helpful to be reminded of them here.

1.4.1 Definition of the population census

1-31 A population census is the total process of planning, collecting, compiling, evaluating, disseminating and analysing demographic, economic and social data at the smallest geographical level pertaining, at a specified time, to all persons in a country or in a well-delimited part of a country.

1-32 A description of the population of any country is basic to the production and distribution of its material wealth. In order to plan for, and implement, economic and social development, administrative activity or scientific research, it is necessary to have reliable and detailed data on the size, distribution and composition of population.

1-33 The population census is a primary source of these basic benchmark statistics, covering not only the settled population but also homeless persons and nomadic groups. Data from population censuses should allow presentation and analysis in terms of statistics on persons and households and for a wide variety of geographical units, ranging from the country as a whole, sub-national and local administrative area, down to individual small localities or city blocks.

1.4.2 Definition of the housing census

1-34 A housing census is (similarly) the total process of planning, collecting, compiling, evaluating, disseminating and analysing statistical data relating to the number and condition of housing units and facilities as available to the households pertaining, at a specified time, to all living quarters and its occupants in a country or in a well-delimited part of a country.

1-35 The census must provide information on the supply of housing units together with information on the structural characteristics and facilities that have a bearing upon the maintenance of privacy and health and the development of normal family living conditions. Sufficient demographic, social and economic data concerning the occupants must be collected to furnish a description of the prevailing housing conditions and also to provide basic data for analysing the causes of housing deficiencies and for studying the possibilities for remedial action. In this connection, data obtained as part of the population census, including data on homeless persons, are often used in the presentation and analysis of the results of the housing census, if both operations are conducted together or there is a link between them.

1.4.3 Essential features of the census

1-36 The essential features of a population and housing census were originally defined by the International Conference of Statisticians as early as 1853 in Brussels, and by incorporating these features, countries have been able to carry out censuses that have been internationally comparable in terms of methodology and quality over time². Nowadays, these five essential features have been redefined with the aim of ensuring the coherence of census data gathered in different countries with varying levels of technical development and different cultures. Adopting these features – regardless of the methodology of data collection – enables National Statistical Offices (NSOs) to collect population data of internationally comparable quality.

1-37 The five essential features of population and housing censuses are¹:

- individual enumeration;
- universality within a defined territory;
- simultaneity;
- defined periodicity; and
- the capacity to produce small area statistics.

1.4.3.1 Individual enumeration

1-38 The term 'census' implies that each individual and each set of living quarters is enumerated separately and that their respective characteristics are separately recorded. Only by this procedure can the data on the various characteristics be cross-classified. The requirement of individual enumeration can be met by means of:

- the collection of information through a questionnaire-based field enumeration;
- the use of information contained in an appropriate administrative register or set of registers; or
- a combination of these methods.

1.4.3.2 Universality within a defined territory

1-39 The census should cover a precisely defined territory (for example, the entire country or a well-delimited part of it). The population census should include every person present and/or residing within its scope, depending on the type of population count required. The housing census should include every set of living quarters irrespective of type. This does not preclude the use of sampling techniques for obtaining data on specified characteristics, provided that the sample design is consistent with the size of the areas for which the data are to be tabulated and the degree of detail in the cross-tabulations to be made.

1.4.3.3 Simultaneity

1-40 Each person and each set of living quarters should be enumerated at the same specifically-defined point in time and the data collected should refer to a clearly-defined reference period. The time-reference period need not, however, be the same for all of the data collected. For most of the data, it will be the day of the census; but for some characteristics that are not specifically time dependent - such as employment or migration status - it may be a defined period prior to the census.

1.4.3.4 Defined periodicity (or regularity)

1-41 Censuses should be taken at regular intervals so that comparable information is made available in a fixed sequence. A series of periodic censuses makes it possible to appraise the past, accurately describe the present, and estimate the future. It is recommended that a national census be taken at least every 10 years, though some countries may find it necessary to carry out

censuses more frequently because of the rapidity of major changes in their population size and structure and/or its housing conditions.

1-42 The census data of any one country are of greater value nationally, regionally and internationally if they can be compared with the results of censuses of other countries that were taken at approximately the same time. Therefore, the UN recommends that countries should attempt to undertake a decennial census in years ending in “0” or at a time as near to those years as possible. It is recognised, however, that legal, administrative, financial and other considerations often make it impracticable for a country to adhere to a standard international pattern in the timing of its census. In fixing a census date, therefore, such national factors should always be given greater weight than the desirability of international simultaneity.

1.4.3.5 Capacity to produce small area statistics

1-43 The census should produce data on the number and characteristics of the population and housing units down to the lowest appropriate geographical level that is compatible with national circumstance, and for small population groups, while at all times protecting confidentiality of the information relating to individual persons and households.

1-44 Such ‘small areas’ may in some cases be the lowest level of the country’s administrative geography, or are areas created specifically for the census. Use of a grid square geography, which is being increasingly adopted by some countries, offers both a greater level of international comparability at the small area level and greater potential for the mapping of census data.

1.5 Data collection methodologies

1-45 Since the 2000 round of international censuses, the methodology for collecting census information has changed dramatically in many countries throughout the world, but particularly in Europe. As noted at section 1.3 many have moved away from the traditional wholly door-step collection by enumerators and, instead, now adopt either a completely or partly register-based approach in an attempt to reduce both cost and the burden on the public, while at the same time producing more regular and up-to-date statistical information. This section of the handbook describes the several types of data

collection methods that are currently adopted throughout the world, noting the advantages that each method offers and the challenges that they may present

1-46 The three basic data collection methods are:

- *the traditional census*, typically exemplified by the collection of data directly from respondents in the field by enumerators and/or through electronic/online media;
- *the register-based census*, in which all information is collected from several and independent administrative registers and then linked to create a single statistical database; and
- *the combined approach*, in which elements of both the traditional and register-based approaches are combined to an extent that suits national circumstances.

1-47 These methods (and some variations adopted in particular countries) are described in some detail in the UN's *Principles and Recommendations for Population and Housing Census revision 3*¹, but it is helpful to summarise the key points in the following sub-sections.

1.5.1 The traditional census

1-48 The **traditional census** approach usually involves a full field enumeration. It comprises a complex operation of actively collecting information on a range of topics from all individuals and households throughout the country (thereby meeting the requirements of universality) or in well-delimited part of the country. It is taken everywhere over a specified limited period of time immediately near to a given reference date (census day) in order to fulfil the requirement of simultaneity. Data are generally recorded on census questionnaires, being either in paper or, increasingly, electronic format, or via a secure online service provision.

1-49 There are two major methods of enumeration:

- where the information is recorded directly by an enumerator during a doorstep interview with the household; or

- where the household members complete the questionnaire on their own and return it either to the enumerator, or directly to the NSO via a postal service, or online.

In either case, either a single long form is universally used (or a set of separate forms for individuals, households and dwellings) which contain all the questions on which information is required, or a combination of short and long forms used.

1-50 In the latter case, the short form contains only a selected set of key questions intended for universal coverage, while the long form is used to collect information on the full range of census questions (and covering particularly complex topics such as fertility) but from only a sample of households and dwellings in order to reduce the burden on the public and thereby minimise any potential adverse effects on response rates.

1-51 As various methods can be used for collecting the data - including a mailed back or dropped-off questionnaire, the telephone, the Internet, personal visit follow-up, or a combination of such methods - countries employing the traditional design may utilize a multi-mode data collection approach to suit different population groups and circumstances.

1.5.1.1 Necessary conditions for a traditional census

1-52 Questionnaires can be distributed either by post or by enumerators. In the 'more traditional' censuses, addresses are pre-listed by enumerators prior to the enumeration, but many traditional censuses now use either pre-existing national address lists, or a register that is purpose-built for the census. The method adopted will depend on whether the census address list is based on an official postal address list or on the NSO's own address database. It may also depend on the quality of the list(s); coverage may be weak or even entirely missing in some parts of the country.

1-53 A postal service may be used to distribute the census forms but only when a comprehensive up-to-date and nationally agreed list of addresses is available or can be prepared. In traditional censuses, a big problem used to be how to cover uniquely all addresses in the country since many types of address list may have been created for different purposes. For this reason, it is necessary, in some cases, to undertake a field verification of address lists before the census. However, this is costly.

1.5.1.2 Advantages and challenges of the traditional design

1-54 The key advantages of a traditional census are:

- the comprehensiveness of its coverage (thereby fulfilling the criterion of universality);
- the fact that all the data are collected at, and refer to, the same time frame (simultaneity);
- the flexibility in deciding the topics to be covered and design of the questionnaire to meet ever-changing user requirements for data (relevance).
- the potential availability of multi-variate data for local geographic areas and population sub-groups (capacity to produce small area statistics)

Moreover, the focused and time-bound nature of the field operation implies that the data collection is finished in a short period and does not require long-term or constant monitoring.

1-55 A long-standing familiarity with the traditional census means that there is, within the NSO, generally a good understanding of the methodology that requires less cultural training among its staff. Furthermore, the census frame can then become the base for subsequent sampling frames for the NSO's programme of regular and ad hoc surveys.

1-56 However, the biggest disadvantage of a traditional census is its administrative complexity and associated cost. Traditional censuses are widely recognised as the most elaborate, complex and costly data collection activity that a national statistical office undertakes. A significant cost is incurred in providing the training necessary for temporarily employed field staff to attain sufficient levels of expertise in administering the census questionnaires.

1-57 But in addition to the cost, this complex task requires full awareness and cooperation of the public to participate in it. Ever-falling response rates (due both to the burden on the public and an increasingly less compliant society in general) is a key factor in many countries' decisions to move to alternative way of collecting census data.

1-58 Moreover, because of their complexity and expense, such censuses are usually carried out only once every ten years- or five years for few countries. This, together with the relatively long data processing time required (compared to the relatively short data collection period) means that even the most recent census data available can be several years out of date- a significant progress is

made in recent censuses when census data fully captured electronically during the field enumeration (thereby creating a challenge in fulfilling the criterion of timeliness)'.

1-59 Finally, many countries are experiencing increasing difficulties in traditionally enumerating specific population groups, such as those persons with high mobility or with multiple residences, or who are hard to contact for other reasons, such as the very elderly, inner city dwellers, students, the disabled and those with language difficulties (such as migrants, refugees, and other cultural/racial/ethnic/religious minority groups).

1.5.1.4 Variants of the 'traditional' design

1-60 Two methodological variants of the 'traditional' design should be noted.

Annual updates

1-61 The first methodological variant – as adopted in the United States - focuses on counting the population and collecting only the basic mainly demographic data in the census year. A large household survey then collects and tabulates detailed demographic, social, economic, and housing data every year throughout the decade, replacing the need for a census long form to collect this detailed data from a sample of the population. The survey samples a percentage of addresses each year to approximate a long form sampling rate over a certain period of the census cycle (five years in the United States). To achieve increased reliability and quality of the estimates for small administrative/geographical areas, larger proportions of addresses need to be sampled. A fuller description of this method is set out in the UNECE's *Recommendations for the 2020 Censuses of Population and Housing*³.

1-62 The primary advantage of this approach is to provide more frequent and relevant data on population and housing than would be available when a census is conducted only once a decade. In a traditional census design, even when detailed data are released as soon as possible after the census year, data users are required to work with results that are, on the average throughout the decade, seven years old. The production of more timely data to support decision-making at all levels of government is a major motivation for this approach, although the essential feature of simultaneity is to some extent lost.

1-63 However, such a programme might be costly and technically difficult, as it requires a multi-year round of comprehensive planning, development and testing.

Rolling census

1-64 The second variant - referred to as a 'rolling census;' as adopted in France - represents a further alternative approach to the traditional model by conducting a cumulative continuous sample survey, covering the whole country over an extended period of time rather than a single enumeration carried out simultaneously in all areas relating to a specific reference date.

1-65 There are two main parameters that affect the design of a rolling census:

- the length of the period of time of the enumeration, which is linked to the frequency of updates required; and
- the sample rate, which depends on the budget and the geographical levels required for dissemination (country, regions, towns, local areas, etc.).

For example, it is possible to build a sample framework in order to produce: national results with a single annual survey; regional results by cumulating a few consecutive annual surveys; and small area results by cumulating a more substantial number of years' data (five in the case of France). The annual survey may be conducted over the course of a year, or in a particular month or shorter time frame.

1-66 The conditions necessary for the successful implementation of a rolling census methodology depend on the complexity of the sample framework. If the sampling units are addresses, a master address file is a necessary prerequisite. But if the sampling unit is larger, for example at the municipality level, it is only necessary to have enough information to spread the municipalities over the different years as each will be representative.

1-67 The main advantage of the rolling census approach is the greater frequency of updating data: a traditional census can provide only decennial, or sometimes quinquennial, benchmarks, whereas the rolling census can provide annual updates. Furthermore, it offers the possibility of:

- improving the census process year by year;
- introducing new topics as and when they become relevant;
- adopting new technologies as they emerge; and
- the deployment of permanent teams that can focus on the continuous evaluation of data quality and the training of field staff.

1-68 A major disadvantage, however, is that the essential feature of simultaneity is lost in that it no longer provides a snapshot of the whole population, at any one time, thus complicating comparisons between areas due to different enumeration periods. However, a range of mathematical techniques (for example, averaging and/or projections and/or interpolation) may be employed so that the data are a statistical depiction of the average situation at any one specific point in time.

1-69 Also, as the rolling census is only able to cover the whole country over an extended period of time, any respondents that move may be surveyed several times, while some others will not be captured by the census at all, resulting in the potential loss of another of the essential features of the census – universality - unless careful methodological adjustments are made. For these reasons, the rolling census demands a sophisticated methodological design, but one that has to be readily comprehensible and acceptable to users. Indeed, a major challenge of such a methodology is the necessary to explain to the users of traditional census data what the consequences of moving to a rolling approach will be and how they should interpret the data. It is important that users understand the methodology (as is the case, of course, with any census design) so that they will continue to have confidence in the data.

1.5.2 Register-based census

1-70 A totally different approach from the traditional data collection methodology is the **register-based census**. The concept of producing census-type results based on registers developed in the 2000 round of censuses, although it had been debated and tested to various degrees in the Nordic countries since the 1970s. Denmark was the world's first country to conduct a fully register-based population and housing census in 1981, and several other Scandinavian countries succeeded in using this approach to generate census data in the 1990 round of censuses.

1-71 In this basic approach there is no field enumeration element (either full or sample) at all, and the data to be collected is taken by the NSO solely from information on individuals, households, dwellings and building held on administrative registers (usually under the control of other government departments or agencies). Data held on such registers are accessed by the NSO

and then linked at the unit record level to create a statistical database from which the set of census-type cross-tabulations can be produced.

1-72 As noted a paragraph 22 above, administrative registers are produced on the basis of administrative processes to record and store information on units (persons and dwellings) and variables which are defined by administrative purposes for which the registers are maintained. Although the content and process of registers will differ from one country to another, the broad types of the registers are usually very similar.

1-73 All persons within the defined territory who meet the register's rules are enumerated. In concept, the enumeration is taken from a basic **population register** in which the fields for different census attributes are populated from the subsidiary registers relating to specific topics. Some of the many subsidiary administrative sources that can be used for deriving census-type data are:

- Buildings and dwelling or address register
- Business or enterprise register
- Taxation register
- Employment/unemployment (job-seeker) register
- Pension register
- Social security or welfare register
- Student register

1-74 Information is extracted from the registers as it reflects the situation of individuals at the pre-defined census reference date. The timing of the census extraction may require careful thought where register update cycles vary. Registration delays and administrative delays in updating between regional and national databases can otherwise have a serious impact on the quality of the output.

1.5.2.1 Necessary conditions for using administrative registers

1-75 The 3rd revision of the UN's *Principles and Recommendations on Population and Housing Censuses*¹ notes that among the key pre-conditions to conducting a register-based census that the country should have are:

- a national legislation providing for the creation of a population register and authority to use the data contained in it for statistical purposes;
- an established central population register;

- high-quality data in the population register with comprehensive geographic coverage;
- an effective system of continuous updating of the population register;
- harmonized concepts and definitions used in the various registers used;
- a universal personal identification (unique identity) system to facilitate the effective linking of data;
- quality and consistency checks to verify the suitability of the data contained in various registers.

1-76 A more detailed discussion of the range of registers from which census-type data are commonly derived and the necessary conditions for establishing a register-based census methodology are given in Chapter 2. Quality issues will be discussed in Chapter 5. However, some of the advantages and challenges are summarised here.

1.5.2.2 Advantages of using administrative registers

1-77 The primary advantages of a wholly register-based approach are reduced costs, a greater frequency of data and eliminating the respondent burden. Such reductions in costs result from the absence of any field enumeration (usually the most costly and labour-intensive component in a traditional census) and from the need, for any particular census, to process only those data items that have changed since the previous census. Many personal characteristics such as sex, country of birth, citizenship, religion, levels of completed education and qualifications, and housing characteristics such as period of construction, floor space and number of rooms change quite seldomly (if at all in some cases).

1-78 However, it should be recognised at the outset that establishing and maintaining administrative registers involve higher costs than the census alone may justify. The need for the register will largely be based on its contribution to more reliable and efficient administration. The use for statistics may be valuable but is likely to be a secondary consideration.

1-79 Register-based statistics can be obtained for all geographical areas, since registers aim to cover the target population in its entirety, and because detailed geographical information can be obtained for all geographic units, municipalities, and map grids of different sizes. Moreover, register-based statistics are generally available every year. Ever-increasing demands for information create new pressures to step up the production of regional

statistics, but regional data from a decennial population census may not be sufficiently up-to-date to satisfy these needs.

1.5.2.3 Challenges when using administrative registers

1-80 The use of administrative data sources, however, involves certain drawbacks that need to be taken into account. One such disadvantage is the fact that register-based characteristics have to rely exclusively on the information held for the administrative purpose of the available register; this is invariably a non-statistical purpose. This imposes some restrictions with respect to the variables that are available for analysis and may also undermine international comparability.

1-81 Moreover, when a registered data item is changed, new or updated information is not always recorded immediately. In certain cases, such information may not be registered at all. Where this occurs, the register information will not accurately reflect real circumstances nor satisfy the NSO's requirements.

1-82 Furthermore, without a traditional census questionnaire, there is no longer any collection tool for meeting ad hoc needs. In many countries the population census system is an important vehicle of data collection that is used to meet emerging information needs to reflect changing social conditions (such as measuring economic migration or the impact of a pandemic). This flexibility is lost when data are no longer collected by means of traditional questionnaires.

1-83 However, where such deficiencies arise some countries can choose to combine register-based data with information on some variables collected from pre-existing surveys or other sources. Such a 'combined' approach is discussed in the next section.

1-84 The impact of introducing a register-based system on the essential features of the census described in section 1.4.3 above will be discussed in Chapter 2.

1.5.3 Combined approach

1-85 There are some standard data items that may have to be dropped from a wholly register-based census system (as described above) either because the quality of the administrative registers is not sufficiently good enough (at least

for use for statistical purposes) or the information for some census topics is simply not available in any administrative register. In the latter case such topics may include, for example, household composition characteristics, ethnicity, religion, language, occupation and mode of transport to work. An option would then be to collect such information from other sources such as a reduced field enumeration or sample surveys, but then the census becomes no longer wholly register-based, but instead adopts the so-called 'combined approach' referred to earlier.

1-86 Essentially, the combined methodology makes use of registers relevant to a census, complemented by surveys and/or a complete enumeration. The use of survey/enumeration data is intended to:

- improve the accuracy of the population counts;
- provide information for census variables that cannot be reliably based on administrative data;
- check, update and improve the quality of census data derived from administrative sources;
- add additional variables to the census as and when demanded; and
- provide a linking frame in order to bring together different sources.

1-87 The move to a combined-census approach is gaining popularity with many countries. For example, in a review of practices in the 2010 round of censuses in 50 countries in the UNECE region⁴, ten reported adopting a combined approach compared with just three in the previous 2000 round. When asked about their plans for the 2020 round, some 12 countries reported their intention to carry out a combined census as the number reporting a traditional census fell to 22 from a level 40 in 2000.

1.5.3.1 Methods using different field enumeration strategies

1-88 Information on individuals, households and dwellings is collected by combining data from registers with data collected from one or more surveys. Data collection may be based on (a) full field enumeration, (b) an ad hoc sample or (c) existing survey methods. Data from registers are employed not only as a sampling frame or to support field operations, but directly as the data source for some census information.

1-89 The option of a register-based census with sample surveys or full-field enumeration should only be adopted if all the necessary census information is

available from one or more administrative sources and necessary to collect other information in the field, *and* it is possible to link the information from the different sources at the record level. During the process of integrating individual records, care should be taken to check the accuracy of the data and remove any inconsistencies and/or duplications prior to the production of statistical outputs. Quality assessment issues will be discussed in detail in Chapter 5.

1-90 (a) **In those censuses where registers are used together with a full field enumeration**, data from the registers may be pre-filled in the questionnaires, and respondents may be asked to check, update and confirm their details.

1-91 Variables not available from administrative registers can be obtained through the field operation, as in a traditional census. The questionnaire used for this purpose need then contain fewer questions compared to a traditional census form, but still cover the whole population of individuals, households and dwellings. In the 2010 round of censuses in Europe, countries that adopted this approach included Estonia, Latvia, Liechtenstein and Lithuania⁴; Indonesia's 2020 census is another example. Over time, more countries may decide to adapt this model, increasing the use of integrated administrative registers in successive censuses and reducing the number of questions in the questionnaire still further.

1-92 (b) **A variation of the combined approach involves the use of an ad-hoc sample survey** instead of a full enumeration. This option is (like the combined census with full field enumeration described above) open to those countries that do not have all census information available in its registers. In the 2010 round of censuses, countries that applied this approach included Germany, Israel, Poland, Spain, Switzerland and Turkey⁴. The base register can then be used as a sample frame and also, as before, to prefill some information as is the case with a full enumeration described above.

1-93 In such censuses, the ad-hoc sample survey questionnaire can be specifically designed to complete and statistically correct the data derived from registers – covering those variables not available from the register. The sample can be sized and stratified in such a way that data are available for small groups and geographical areas.

1-94 If this option is chosen, some census data can be taken directly from the relevant register, while for other census outputs, information from sample field data have to be weighted to the population totals. In South Korea, for example,

the sample size is 20 per cent. This is, however, only one way to merge register and survey data to obtain census-type results. Israel, for example, uses a population register as the basis for a 100 per cent enumeration, and undertakes a census sample survey to evaluate its accuracy.

1-95 (c) The Netherlands model represent another variant of the combined approach. There, the census is carried out **by combining data from population and integrated administrative registers with existing sample survey(s)**. Different data sources are integrated as part of what is termed a 'virtual census' process. The data for the census exercise are derived from many types of registers and surveys, covering different population groups throughout the country and its sub-regions. Compared to conventional census methods, this process is lower in cost and staff requirements, and uses more frequently updated data sources. More significantly, by combining data that are already available from other sources, the virtual census makes no additional respondent burden, thereby increasing its public acceptability.

1-96 When registers are used along with sample surveys, some census outputs may be produced entirely from the information available from the register(s), while for other census tables, information from the survey(s) duly weighted to the population totals, can be used. The survey-based data can also serve as an independent source to evaluate the accuracy of the register counts.

1.5.3.2 Necessary conditions

1-97 The necessary conditions for implementing such a combined census are broadly similar which ever type of approach is adopted. Any one option can only be adopted if all necessary census information is available from one or other of the sources, *and* it is possible to link the information from all the different sources at the record level.

1-98 The data sources should include verified and accurate personal information (name, ID number, date of birth, sex, marital status, family structure, etc) and a dwelling register. In an ideal situation, a centralized 'base' register can be envisaged to include unified identity codes for both people and address components in order to link more efficiently the related register and survey data.

1-99 As with a wholly register-based approach the link between persons and their dwellings is equally important in order to identify the household unit. All

the other conditions relevant to a register-based census discussed above will apply to the register components of a combined census.

1.5.3.3 Advantages

1-100 A key advantage of the combined census with a full field enumeration is that the population counts in this approach may be more precise than in a wholly traditional census because of the generally more up-to-date information contained in the population register that helps to minimise coverage error. It may also be more accurate than in an exclusively register-based census because of the checking against the actual reality that a complete enumeration provides that can then correct for any accumulated errors in a population register.

1-101 Information not available from a combination of data taken from registers is obtained from a traditional field operation approach, thus permitting maximum geographical and statistical detail.

1-102 The longitudinal perspective that the use of registers provides is also present in this census approach, for example, by using the population register as the frame.

1-103 The advantages of a combined census with sample field data are that it is much less costly than any census with an element of full field enumeration and imposes less of a burden on respondents (and enumerators). A traditional census can face many privacy objections to the seemingly intrusive collection of personal (and sometimes potentially sensitive) information. This increases the extent of non-response. There are generally fewer objections, and less risk of non-response, to a combined census in which data are collected only from a sample of the population. It may be the case that data from such a combined census will provide sufficiently reliable results, because they are derived to some extent at least from registers covering the whole population.

1.5.3.4 Challenges

1-104 The combined approach with a full enumeration is much more expensive than a wholly register-based censuses, because of the cost of the field enumeration component of the data. However, the field costs will generally be less than traditional census since the knowledge of the location where every person is registered facilitates the use of more efficient collection methods.

Where only a sample field enumeration is carried out the cost will be proportionately less than for any census with an element of a full enumeration.

1-105 Response burden, other factors being equal, also falls somewhere between the minimum imposed in those censuses without any specific data collection operation and the maximum for censuses with a full enumeration and no administrative information support. Where only a sample field enumeration is conducted through ad hoc surveys, public awareness of the census and their obligation to participate is significantly reduced.

1-106 A further disadvantage of the ad hoc survey approach is that this option requires more processing to produce the tables from the unit record information as weighting problems may arise.

1-107 The use of pre-existing surveys also presents a number of problems compared to the use of ad hoc surveys. The timing, content, statistical definitions and sampling approaches used in an existing survey may not be appropriate to allow the data to be readily combined with data from the administrative source(s) resulting in poor coherence between the different data sources. For example, many major household surveys are not designed to cover persons living in institutional households (such as student accommodation, hospitals and nurses living accommodation, prisons, and military establishments), thereby requiring additional sources of information for such residents.

1-108 A weakness of the use of sample data is that for the variables derived from the surveys the sample size may not be sufficient to offer the geographically detailed outputs that are an essential feature of the census. Importantly though, the use of data from existing regular surveys will often mean that time series of data are available. By combining samples for several survey exercises, it may be possible to produce reliable estimates for small geographical areas.

1.6 Current profile of global censuses

1-109 In its role as the Secretariat to the 2020 World Programme, the *United Nations Statistics Division (UNSD)* has undertaken a variety of activities aimed at ensuring that countries undertake a census during the 2020 round, covering the period 2015–2024. It prepared the methodological framework for 2020 round population and housing censuses that was anchored by the revised *Principles*

*and Recommendations for Population and Housing Censuses*¹ issues in 2015. This has been supported by a set of additional handbooks on:

- population and housing census management⁶;
- the use of contemporary technologies for population and housing census data collection⁷;
- the use of censuses in measuring international migration⁸; and
- population and housing census editing, including real-time editing⁹.

1-110 As part of its responsibility to monitor and regularly report to the UN Statistical Commission on the implementation of the 2020 World Programme, UNSD developed and launched a short, 11-question survey sent to all NSOs in July 2019, the purpose of which was to collect information on countries' plans and outcomes of the 2020 round of censuses.

1-111 The results of the survey¹⁰ show that the traditional census, based on full field enumeration, is still the predominant methodology in conducting population and housing censuses (see Table 1). Out of 158 countries that responded to the survey, 112 countries (71 per cent) reported that they had used, or will use, the traditional approach to data collection in the 2020 round.

Table 1- Main census methods for 2020 round population and housing censuses

	All 158 responding countries	Africa (38)	Asia (36)	Europe (41)	Latin America and the Caribbean (28)	Northern America (4)	Oceania (11)
<i>Number</i>							
Traditional census	112	37	25	12	28	1	9
Fully register based census	16	0	3	12	0	1	0
Combined census - Registers and full field enumeration	21	1	5	11	0	2	2
Combined census - Registers and sample survey	8	0	3	5	0	0	0
Rolling census	1	0	0	1	0	0	0
Other	0	0	0	0	0	0	0
<i>Percentage</i>							
Traditional census	71	97	69	29	100	25	82
Fully register based census	10	0	8	29	0	25	0
Combined census - Registers and full field enumeration	13	3	14	27	0	50	18
Combined census - Registers and sample survey	5	0	8	12	0	0	0
Rolling census	1	0	0	2	0	0	0
Other	0	0	0	0	0	0	0

1-112 Only 16 countries (10 per cent) reported carrying out a full register-based census, though a slightly larger number (29 countries, 18 per cent) have adopted or are planning to adopt a combined approach by either carrying out full field enumeration (13 per cent) or by conducting sample surveys (5 per cent) to collect those data on census topics that are not available in the registers.

1-113 It should be noted, however, that this apportionment of register-based and combined censuses is not global. The majority of countries that are using administrative registers to a greater or lesser degree in the 2020 round (28 out of the 45) are in Europe and 11 of the others are in Asia. According to the survey results, although several countries throughout Africa, Latin America or the Caribbean were planning to use administrative registers in the current round, no country was able to do so at the time of this draft. They may be planning to change their census methodology in the next round. Moreover, in Europe the use of administrative registers - either in a wholly register-based methodology or in a combined approach is - now becoming common methods rather than the exception, with more than two thirds of responding European countries reporting these data collection methods in their 2020 round of censuses according to the UNSD survey.

1-114 This change of approach in Europe has been developing since the 2000 round where the methods adopted for collecting information in population and housing censuses have been evolving continuously over the past decades in order to:

- adapt to changes in society especially in public attitudes and behaviour and in greater levels of mobility;
- meet the ever-demanding requirements of data users;
- capitalize on opportunities offered by technological innovations and development, and statistical modelling techniques
- take advantage of new or different wide-ranging data sources such as registers and administrative files and Big Data; and
- reduce costs.

1-115 Moreover, the unprecedented circumstances of the global Covid pandemic in 2020/21 have no doubt accelerated this transformation in the method of collecting census information that was already underway in many countries. The days of waiting a year or two for cleaned, quality assured, tabulated census data to be made publicly available may be coming to an end, as researchers, policymakers and other census users demand, and come to expect, short-time data availability with, perhaps, less insistence on 100 per cent accuracy.

1-116 This move away from the traditional census particularly in UNECE countries in the first decades of this century is evident from the profile of the censuses reported by member states in a number of UNECE surveys on the 2000-2030 rounds (see for example UNECE (2008)¹¹, UNECE (2014)¹² and UNECE (2016)¹³).

1-117 The most recent of these surveys was carried out in September 2018 and was reported in a paper presented to the Conference of European Statisticians by the Steering Group on Population and Housing Censuses and which was published in August 2020 under the title *New Frontiers for Censuses Beyond 2020*¹⁴.

1-118 As can be seen in Table 1.1 some 35 countries (two thirds of those in the UNECE region) conducted traditional census in the 2010 round. This was, at that time therefore, still the preferred general methodological approach throughout the region, but was less so than was in the 2000 round in which over four fifths of countries adopted this approach.

Table 2 Type of census, 2000-2030 rounds, UNECE countries

Type of census	2000	2010	2020	2030
Traditional census*	40	35	14	2
Combined census	5	10	17	14
Register-based census	3	9	13	23

* Including the Rolling Census in France

1-119 Among the other countries in the 2010 round, 10 carried out a census combining the collection of data through field enumeration with use of administrative register - this was double the number that did so ten years earlier. And the number of countries that had moved to a wholly register-based system tripled from 3 to 9.

1-120 Ten years on, the continued decline in the traditional census is evident. Only 14 countries are conducting (or have conducted) a traditional census in the 2020 round (just of a quarter of the countries that responded to the 2018 UNECE survey) and almost the same number have moved to a wholly register-based approach. But now the largest proportion of countries (over two thirds) are adopting a combined census.

1-121 By 2030 our world may look very different – and not only because of the coronavirus pandemic. The ability to collect information from a wider variety of sources, more rapidly, more efficiently, and with minimal direct public involvement, is increasing all the time. Knowing this, citizens may rightly wonder why so much time, effort and money should be spent on asking them to complete a questionnaire - often the same questionnaire - every ten years.

1-122 The broad overview is that for the 2030 round, 23 countries report that they plan to undertake a wholly register-based census. This represents a half of all the countries that responded to the survey and shows a clear paradigm shift in data collection methodology.

1-123 In addition to the 13 countries where the use of administrative data as the almost exclusive source of census information is now fully established, a further ten countries plan to adopt a register-based census in the 2030 round, including two who plan to move directly from a traditional census culture in one go.

However, the majority of countries who plan to move to a register-based census in the next round will do so from a position of having already laid the foundations with the use of administrative data to a varying degree in the 2020 round.

2 Moving to a register-based approach

2.1 Introduction

2-1 As discussed in Chapter 1, the adoption of a register-based census methodology offers a number of potential advantages to an NSO including decreased cost. While the COVID 19 pandemic may have provided further benefit (indeed in some cases a necessity) for some NSOs to consider the adoption of a register or combined census methodology, strategically the consideration of such a move should be set within the context of a wider discussion on the use of unit record administrative data in their statistical systems more generally.

2-2 However, any such move needs to be carefully managed, and there are necessary conditions relating to data, technology, legal and stakeholder issues and human resources that must be considered for a successful transition from a traditional to a combined or register-based census. This transition may therefore give rise to some challenges. For example, although many NSOs may have a long history of aggregating data taken from administrative sources for producing vital statistics that are derived from a vital registration system, they may have only limited or no experience of using unit level administrative records for other statistical purposes.

2-3 Such challenges are set out in section 2.4 and can present significant obstacles for some countries. But firstly before these are discussed, section 2.2 examines those pre-conditions that have to be in place before *any* register-based system can be planned (such as a legal framework, stakeholder and public support, access to appropriate registers, data protection methodology and technical expertise). The particular circumstances that would enable a *specific* register-based census to be conducted are then discussed in section 2.3. These include specific census legislation, knowledge of users' requirements, and the ability to create population and address registers, and access to the relevant administrative registers, including specialized registers.

2-4 The chapter concludes with a brief exposition of the effect that moving to a register-based statistical system has on the essential features that define a census.

2.2 Necessary pre-conditions

2-5 This section of the handbook examines those pre-conditions that have to be in place before *any* register-based statistical system can be planned and implemented.

2.2.1 NSO's long-term project vision for a register-based statistical system

2-6 In many countries, the move to a register-based (or combined) census is a core part of the strategy for the NSO to develop or adopt a register-based statistical system. This strategy will transform the statistical production system, and so represents a paradigm shift for all functional aspects of the NSO. Indeed, it may often be the case therefore that the transition to a register-based census is only the first step in the NSO's long-term vision for a more widely register-based statistical system. The initial cost and resource required to develop and implement such a system for a census is only likely to be fully justified once it has been applied to the NSO's other core activities.

2-7 Indeed, one of the United Nations Fundamental Principles of Official Statistics¹⁷ concerns cost-effectiveness and suggests (at Principle 5) that *“Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records”* taking due regard to quality, timeliness, costs and the burden on respondents. An explanatory note to Principle 5 also stresses the importance of cost-effectiveness, and goes on to say that: *“Statistical offices must be cost-effective, making the best choice of concepts, sources and methods by balancing quality, timeliness, costs and the reporting load of respondents The overall cost-effectiveness of an agency is influenced by organizational planning and operation, the sound application of statistical methodology, exploitation of information and communication technology and also access to administrative records”*.

2-8 As UNECE has noted¹⁵ a pure register-based statistical system could be defined as one in which all statistics (for a particular domain or set of domains) are produced exclusively from administrative sources that have been combined into two or more linked statistical registers. In practice, such a purely register-based statistical system is relatively rare, as small-scale statistical surveys are often needed for quality assessment or to overcome coverage issues for specific variables or sectors of the population. A more pragmatic approach is therefore to use the term ‘register-based statistical system’ to refer to a system based

primarily on administrative data that have been organized into linked statistical registers.

2-9 However, such a strategic move requires a number of pre-conditions to be in place. These pre-conditions (described below) can also be seen as the fundamental building blocks for the register-based system. Without these, it is not feasible to consider a register-based census in the short-term option. However, it may be possible to use the census project to establish the necessary pre-conditions.

2.2.2 Legal framework and role of the NSO to be established

2-10 Whatever type of census an NSO conducts, it requires a legal framework. This is normally set out in a Census Act or is encompassed within the more general provisions of a national Statistics Act. Such legal frameworks for many NSOs have been designed for statistical systems mainly based on traditional censuses and surveys. The legal framework for a register-based system requires a different approach, recognizing that the system has moved to a register-centric model, where administrative data (whether taken from public or commercial agencies) are transformed to meet statistical needs.

2-11 To be able to conduct a register-based or combined census there must be legal provisions that prescribe the access to, and protection of, such administrative data. The NSO must have legal authority to access the relevant administrative data, ideally free of charge, from any relevant public authority sources. To avoid legal uncertainty or dispute, such legislation may need to stipulate any exceptions to such a right of access in circumstances that would risk the protection of public order or national security.

2-12 A further key element is that the legal frameworks should empower the NSO with some influence or authority in the creation, revision or deletion of administrative data in order to ensure that the data to be used by the NSO in carrying out its designated statistical functions are relevant, consistent and reliable.

2-13 In turn, the NSO must be subject to a legal obligation to: use the data only for statistical purposes; protect the confidentiality of the administrative data it obtains; and adhere to a 'one-way traffic' only principle. Essentially, the relevant legislation should specifically prohibit access to the data held in the NSO's statistical registers by any other person or body.

2-14 Such legislation needs to accommodate legal considerations not only for the NSO, but also for the administrative data suppliers, as well as being consistent with wider government frameworks covering privacy/data protection and data security. In addition, some countries may have local frameworks (in the form of specific state laws, for example) as well as wider regional commitments. For example, it may be the case that there are legal restrictions that data can only be used for specific statistical purposes, and that the confidentiality of individual records should be maintained. Thus, there may be specific restrictions on the use of data on unincorporated businesses, particularly sole-proprietorships, where business data could be considered to be personal data relating to the owner of the business. Many (if not all) countries have data protection legislation covering information about individual citizens, therefore it is important to make clear distinctions between what constitutes business and personal data in such cases¹⁵.

2-15 In moving to a register-based system, it may therefore be necessary for the NSO to update or amend its legislation. While the NSO will focus on the specific legislation necessary to enable it specifically to establish a register-based system, it is possible that agencies providing the administrative data may also need to update their own legislation. This includes, for example, ensuring that agency legislation enables and supports the supply of unit records and their associated metadata to the NSO for statistical purposes within the context of specific data protection/security legislation.

2-16 UNECE have produced useful guidance on modernizing statistical legislation¹⁶ to which NSO may wish to refer to ensure that any revisions that are necessary to implement a register-based statistical system are consistent with the Fundamental Principles of Official Statistics, the European Statistics Code of Practice (where appropriate) and the OECD Recommendations on Good Statistical Practice.

2.2.3 Political support

2-17 Political support for moving to a register-based statistical system is essential. This includes support for the overall approach as well as for the necessary legal changes (discussed above) and for removing any barriers to accessing administrative registers.

2-18 Any legislative process required to make provisions to implement a register-based statistical system can take time. While in some countries (such as Norway and other Nordic countries) there has been considerable support for the use of administrative registers for statistical purposes, this is not always the case in those countries where statistics may often be regarded as a relatively low priority by legislators, and where a sustained period of lobbying politicians and highlighting the benefits of using administrative data may be necessary. In the UK, France and Germany for example, historically there has been political and popular opposition to moving to a register-based statistical system. In the case of the census, this means that the NSO needs to ensure that key stakeholders – particularly politicians - are well informed about the impact of the change in methodology on the census process as well the likely impact on census statistics.

2-19 Given all the efforts that are usually needed to introduce or revise statistical legislation, it is therefore necessary to make the most of the opportunity. In particular, it is essential to avoid the mistake of proposing legislation that just meets current requirements. It may be ten or more years until the next opportunity to revise legislation, so it is necessary to have a long-term strategy for the use of administrative data, and to ensure that the legislative proposals meet all envisaged requirements for the foreseeable future. In this way, legislation can be seen as a short-term barrier, but a long-term opportunity.

2.2.4 Public approval

2-20 In many countries, the attitude of the public to data linking and sharing between government agencies is a key factor in determining the extent to which administrative data can be used for statistical purposes. Such public opinion varies considerably from country to country. While the law might give a legal licence for the NSO to conduct a combined or register-based census, public approval is also necessary to ensure that such a census is acceptable. This is often easier said than done and might be more difficult to achieve than establishing the legal base.

2-21 While in some countries people may get the impression that, in the course of a traditional census, ‘big brother is watching you’, elsewhere the use and linkage administrative data collected for non-statistical purposes may be seen to be even more intrusive as the public then has no control at all over the information about them that is to be disclosed and may prefer to answer census

questionnaires directly. This issue is discussed as a particular challenge to overcome at section 2.4.9 below.

2.2.5 Cooperation between NSO and register-holders

2-22 Good cooperation between the NSO and other (mainly government) authorities is vital for using administrative data sources in statistical systems, and for the census in particular. Notwithstanding any legal obligation that may be imposed on government agencies and other authorities to provide the NSO with the requisite data as and when required, the use of registers requires good on-going working relations between the NSO and the administrative register-holders. This will help to engender a better understanding among administrative agencies of the role that their data play in the statistical system, and the importance of keeping the NSO informed about any changes to the administrative unit record data and metadata.

2-23 For example, it is necessary that the NSO should know when the relevant administrative unit records (and any accompanying metadata) can be made available before any register-based statistics can be produced. Particularly for the purposes of a combined or wholly register-based census (and particularly so in the case of the latter) the NSO is heavily dependent on administrative data holders to comply with their agreed or legal obligation to provide good quality data on time. If data holders fail to deliver, it is usually the NSO that is held responsible for the failure to publish census statistics punctually. Only with regular contact between the NSO and the administrative data holders can the success of any register-based statistics be maintained.

2-24 It is vital to inform administrative data holders how important their data are for the NSO and how their data are to be used for statistical purposes. In return, the NSO needs detailed information from data holders on how their respective administrative systems are organised and what (if any) changes are being planned or anticipated. Such collaboration often take place within the frame of a national forum of register keepers but can also be conducted bilaterally.

2-25 In addition to a legal framework and good contacts with other authorities, signing cooperation contracts or service level agreements could help in supporting the NSO's statistical programmes and particularly the census process. In theory, administrative data sources from non-governmental

authorities could also be used in the census, but this often creates particular privacy and data quality concerns and involves commercial considerations; private sector data are more often than not only acquired at a substantial cost to the NSO.

2-26 The necessity for the NSO to establish and maintain good working relationships with register holders for the purposes of assessing the quality of the administrative data is discussed in more detail in Chapter 5.

2.2.6 Organizational frameworks

2-27 Once the legal frameworks to permit the use of administrative data are in place, it is necessary to consider the organisational arrangements necessary to facilitate data flows¹⁵. Typically, these can be set out in a formal written agreement between the NSO and the relevant government departments or agencies in the form of a “service level agreement”, “protocol” or “concordat”.

2-28 There are some key features that should be present in any such agreement, of which the following are the most important:

- Detailed description of data covered: This will include information identifying the data set and the variables contained within it.
- Frequency of data supply: This will specify when and how frequently the administrative organisation will supply data.
- Quality standards: These set the parameters for the quality of the data supplied. Examples might include requirements for addresses to be in certain standard structural format, or for a maximum proportion of missing or erroneous variables, to ensure that the data received are fit for purpose. The priorities assigned to different variables, and hence the effort put into quality assurance, will often differ between administrative and statistical organisations, therefore agreeing common standards can be difficult.
- Provision of metadata: It is important that data flows are accompanied by the relevant metadata, which can include dates, descriptions for any codes used, information on the units used, etc.
- Confidentiality rules: It is important to set out what the data can be used for, what rules and procedures will be in place to prevent disclosure, and in what circumstances the data can be passed on to the NSO’s users.

- Technical standards: Technical frameworks refer to the mechanisms by which data are transferred, as well as any relevant data or metadata standards. Data transfer mechanisms can take any form of a secure electronic link, such as real-time or periodic (weekly or monthly) updates. There are a number of international standards for data and metadata transmission, including XML and SDMX name just two. Some countries also have national versions, particularly for data transfers within government. It is therefore important to agree which standards are to be used.
- Period of agreement: Agreements may be for a fixed period, but if so, they should include provisions to renew or extend them as necessary. An alternative approach is to have an agreement that is considered valid until one of the parties wants to make a change.
- Contingencies for changes in circumstances: It is important for the statistical organisation to have advance warning of changes affecting the administrative source. The agreement should specify that any proposed changes are communicated to the statistical organisation as soon as possible, to allow the impact of the changes on statistical outputs to be minimised.
- Procedures for resolving disputes: The agreement should specify the method to resolve any disputes between the statistical and administrative organisations, which will normally be to escalate issues to senior managers, or possibly even to the relevant ministers.

2.2.7 Knowledge of administrative registers and initial assessment of their quality

2-29 Most countries that are contemplating the move to a register-based statistical system will have an extensive range of administrative records and registers, many of which may be suitable for use in such a system. To determine whether or not they can be used as a data source, it is necessary for the NSO to identify the availability of such registers, and then assess the suitability of each source and the variables contained for the particular purposes required. Section 2.3.4 notes the importance of identifying those registers that are available for specific use in the census, and Chapter 3 discusses these in more detail.

2-30 An initial assessment needs to be undertaken as a pre-requisite to determine what registers exist and are of potential value for statistical purposes.

As the Eurostat Concepts and Definitions database notes, in the context of use for statistical purposes registers should have a number of essential features:

- they should be 'systematic'(that is, regular and organized);
- the data held should be at the 'unit record level'; and
- they should be capable of being updated.

2-31 As will be discussed in Chapter 3, suitable administrative sources should also have a comprehensive coverage of the target populations. The existence of large numbers of unregistered units (for example undocumented immigrants, nomads or businesses operating in the informal economy), will make it extremely difficult to produce meaningful register-based statistics.

2-32 It is also important that data sources are available to the NSO in digital format. This helps in the transfer of data to the NSO, but most importantly is fundamental for the statistical transformation (that will be discussed in more detail in Chapter 4). Where only paper registers are available, it will be necessary to create digital versions. This can be an expensive and time-consuming task, but it will benefit the source agency as well as the NSO.

2-33 The initial assessment will be followed by more in-depth statistical quality assessments of the definitions, concepts, content, reference date, coverage, accuracy, etc.), whether the administrative registers contain identifiers and are available (or could be made available) in digital formats, and how the data will be shared. Furthermore, such assessment of the data sources should also determine whether or not sufficient metadata (data that define and describes the record data) exists that can be provided by the register holders to allow the NSO to understand the register, including source of the data, purpose, contents, creation and updating processes. These issues are discussed in more detail in Chapter 5 within the context of assessing the quality of administrative sources and data for the purpose of the census.

2.2.8 Unified identification system

2-34 To facilitate the linkage of administrative data from different sources that relate to the same unit, it is preferable to have unique ID-numbers at the unit record level that are common across all respective registers. In the absence of such unified systems it is extremely difficult and laborious, if not impossible, to link unit data reliably from different registers (as is noted in the section 2.4

below). More will be said about this in the discussion on the linkage of data sources in Chapter 4).

2.2.9 Institutional infrastructure and capacity

2-35 In order to move to a register-based system, it will be necessary for NSO staff to develop beforehand a full range of appropriate skills. These include relationship management, information management, data mining and analysis and associated subject matter skills. Specialized statistical skills will also be needed to support the data linking processes. While NSOs generally have a long-standing expertise in information management, it is important to recognise that a register-based statistical system will require statistical knowledge and expertise, particularly in the linkage and matching of data in general, as well as specific subject matter knowledge and experience. It is, therefore, necessary for the NSO to develop an infrastructure, and to implement a training programme, that is designed to help staff acquire the requisite statistical skills to support the necessary transformation processes. Knowledge of official statistics systems will be particularly relevant. Capacity is needed to be able to identify those administrative registers that can be used for censuses, to assess their quality in terms of completeness and coverage, and to analyse possible options and design census processes with an appropriate methodology taking into account specific conditions of the national context.

2-36 Ideally, the administrative registers and their variables should have concepts, definitions, classifications, data formats and reference periods, consistent with the NSO's statistical requirements. In practice, this is unlikely (as noted in Section 2.4), and so the NSO will need to prepare systems and processes to transform the administrative data into statistical records.

2-37 It is, therefore, necessary for the NSO to have the requisite statistical skills and infrastructure to support the necessary transformation processes. Such transformations will include:

- coding to standard statistical classifications;
- editing records, including correcting errors in identifiers;
- linking unit record data between registers; and
- creating new variables from data from different sources.

Without these skills and expertise, it may not be possible to produce trusted statistical outputs.

2.3 Census-specific pre-conditions

2-38 In addition to the general pre-conditions that are necessary to instigate a register-based statistical system, a number of specific pre-requisites need to be in place in order for the NSO to be able to conduct a register-based census at any one time. These are discussed in this section

2.3.1 Census-specific legislation

2-39 Notwithstanding the existence of a generic census law or more widely encompassing primary statistics legislation, all countries are likely to require a specific law and/or government regulations relating to each particular census. In some countries, such an existing legal framework may allow the NSO only to conduct a traditional census with a field enumeration. In these cases, it will be necessary to amend the census-specific legislation to ensure that the NSO has a clear legal mandate to use administrative data for the purposes of the census, subject to any other requirements or restrictions set out in other statistical legislation. If not already provided this may need to include the obligations on the relevant administrative agencies to provide the required administrative unit record data and any relevant metadata.

2-40 Depending on the local legislative environment and country practices, such census legislation may specifically prescribe the census reference date, and the topics or variables that will be collected. Where a combined census is being mandated it will be necessary to distinguish between those data items that are to be collected directly from households through the use of a questionnaire and those which are to be derived from administrative sources. Moreover, NSOs conducting combined censuses will also need to ensure that the new or amended legislation provides the authority to link administrative data with data collected using traditional methods.

2-41 As noted at sections 2.2.2 and 2.2.6 it is important to remember that changing legislation and creating organisational frameworks takes time – in some instances a lot of time! This needs to be factored into the census plan. Only in those countries in which a wholly register-based statistical system is well-established (and where the concept of ‘a census’ is fast becoming a fading memory) will there perhaps no longer be a need for any census-specific legislation.

2.3.2 Data user requirements

2-42 All censuses should be based on a clear set of data user requirements, including information needed for planning, policy modelling, defining electoral boundaries, etc. Such requirements will be transformed into a series of census outputs (tables, analyses, maps, etc.). Users typically want each census to provide at least the same level of detail of information as in the previous census. However, this is not always possible when the census methodology changes radically.

2-43 Whenever statistical methodology changes, it is good practice to be transparent and share with users any information on the potential impact. Moving to a register-based statistical system may occur over a long period of time, and have significant implications for users in particular, to the quality and consistency of census outputs over time.

2-44 The views of users are, therefore, important in helping to determine if the benefits (less cost, less response load, improvements in timeliness, improved consistency) balance any discontinuity in outputs and time series due to changes in methods, concepts and definitions.

2-45 It is important to inform and consult stakeholders before any census. The census will not be relevant if the NSO only produces statistics that are based on easily accessible registers. Therefore, it is critical to have a clear set of user requirements before beginning the process of assessing the quality of the administrative data sources. Users will become disappointed if their expectations are not met. Even well-informed users can become very critical if they believe that they are going to lose some of the information that they had access to in the previous census. However, it is usually not possible to satisfy all users, and some disappointment is often unavoidable when adopting a new census methodology. In order, therefore, for the census to be relevant to users and other stakeholders, their requirements should drive the detail and structure of data items to be derived from the registers.

2-46 It is necessary, therefore, to have a communication strategy for stakeholder engagement that should encompass some, or all, of the following goals:

- create a transparent environment concerning the plans of the NSO

- assure users that their requirements will be taken into consideration;
- promote the benefits of using administrative data and demonstrate that the information collected will continue to be kept secure;
- strengthen partnerships with stakeholders so that the NSO can benefit from outside expertise;
- make stakeholders part of a successful transition to a new census approach.

2-47 Openness and the clear identification of new opportunities and benefits for stakeholders will help to gain their acceptance. It is particularly important that there should be adequate consultation on any change in the provision of those statistics that have financial consequences for stakeholders (such as transfers of money to municipalities).

2.3.3 Confidentiality and public approval

2-48 The political decisions concerning the use of administrative data in a census can be highly influenced by public opinion. Engagement with the public is key. In many countries people have become increasingly sensitive towards the collection and analysis of personal data by governmental authorities. (For example, some countries have had much debated the use of data collected in COVID19 tracing operations.) In addition to outlining the general benefits and risks of the use of administrative data, information given to the public should focus on the confidentiality of the linked individual data from different sources.

2-49 The use of administrative data in the census must be consistent with the Sixth Fundamental Principle of Official Statistics, which requires that “the use by statistical authorities of individual data, whether they refer to natural or legal persons, be strictly confidential and used exclusively for statistical purposes”¹⁷. This includes ensuring that individual data collected for statistical purposes is not passed onto other government authorities. As a consequence, it means that the NSO must ensure that none of the edited or linked unit record administrative data is returned or passed back to government agencies or authorities.

2-50 There must also be strict rules around access to the administrative records within the NSO (often set out in the relevant statistical legislation) and equally importantly, strong statistical methodologies to preserve the confidentiality of unit record information in statistical outputs, including where NSOs provide access to anonymised microdata files for researchers.

2-51 In a traditional census, privacy concerns may lead to lower response rates or the deliberate disclosure of incorrect information. It is becoming more and more difficult to correct for such unit and item non-response. So, on the one hand the public might prefer a situation where fewer questions are asked if the equivalent information is already available from administrative sources. Where public opinion generally accepts, or is in favour of, data sharing, the increased use of existing data sources can help to enhance the prestige of the NSO by making it more efficient and cost-effective.

2-52 On the other hand, in a register-based or combined census people may feel uneasy about, or even object to, information from different administrative data sources being reused and linked in a census. It may not be clear to them that in the census the information is anonymised and only used for statistical purposes. If census forms are not required and only registers are used, the public will generally be less aware that a census is being conducted. However, the absence of any public reaction should not be misinterpreted as public approval.

2.3.4 Knowledge of relevant administrative sources

2-53 When a country wants to move from a traditional census towards a combined or register-based census, building up a wide-ranging knowledge of the data held in administrative sources that are both relevant and fit for purpose is important before making the move. This includes assessing possible sources for all the census topics required by users, as well as sources for the base registers discussed at section 2.3.5 below. The assessments will include identifying the availability, accessibility and suitability of different sources (and ultimately each variable within each source).

2-54 Although building up such knowledge should occur early in the census planning, the effort needed to make this process successful should not be underestimated. Many lessons about failures and successes can be learnt from countries conducting combined or register-based censuses, but the national context should always be taken into consideration. It is never advisable for an NSO to simply adopt the methodology of another country when setting up a combined or register-based census. However, by learning from the experiences of others the transformation period can be shortened drastically.

2-55 The process described in section 2.2.7 will be particularly relevant when assessing the viability of using administrative registers for the purposes of the census. Those sources of data from which census-type variables are most commonly accessed are discussed in Chapter 3.

2.3.5 Capacity for assessing the availability and quality of administrative registers

2-56 There is a wide variation between countries in terms of the availability and quality of administrative registers and the capacity of national statistical systems to make best use of administrative data for statistical purposes. Therefore, there is an urgent need to strengthen the capacity of national statistical offices or census agencies, especially those in developing countries, to improve uses of administrative registers for small area statistics. Capacity is needed to identify administrative registers that can be used for censuses, to assess the quality in terms of completeness and coverage of administrative registers and to analyze possible options and design census processes with an appropriate methodology taking into account specific conditions at the national context. It should be noted that a long-term perspective is needed for the development of a register-based census through the development or improvement of key registers in the country.

2.3.6 Availability of base registers for censuses

2-57 It is self-evident that in order to conduct a register-based census, there must be a range of good quality statistical registers available to the NSO (or which the NSO is able to create from the administrative registers). Typically, these registers are categorized as being either base or specialised/supplementary registers.

2-58 Base registers are those that hold the basic information relating to the stock of the entire set of 'population' units that are being counted in the census – typically persons, dwellings and buildings. Base registers are created and then linked at the unit record level to information supplied from supplementary registers (sometimes referred to as specialised or subject-matter registers) from which one or more of the characteristics that may be required for the census can be derived) in order to create the requisite census attribute variables of the population and dwellings from which the census statistics are then produced.

2-59 Chapter 3 discusses base and supplementary registers in more detail, but it may be helpful to summarise here the role of the base registers as a necessary pre-requisite for transferring the census methodology from traditional census to a combined or register-based census.

2-60 The three most common base registers are:

- *Population register* – records all residents in the country, together with their basic demographic information, and updates for births, deaths, changes of marital status, international migration (both into and out of the country), and internal migration (by recording changes of address within the country).
- *Building and/or dwelling register* - usually held by land and property valuation agencies or by local or municipal authorities responsible for the development of housing policies and urban planning. They can provide information on the location of all places where people might live (by distinguishing between wholly residential, partially residential and non-residential buildings) as well as some of the core census housing topics such as ownership, type of housing (for example whether a detached house, terrace house, apartment), whether or not the housing is rented, living area, number of rooms, condition concerning availability of electricity, bathroom, kitchen, and toilet, and year of construction. With respect to a register-based population census such registers are of fundamental importance in the harmonisation with the population register, in that every person that is allocated a dwelling number can be associated with a particular housing unit, and (in some cases) a distinct household.
- Many countries also use an *establishment (or business) register* as a base register for the census. Such registers usually contain information about business establishments (including their industry, institutional sector, size of workforce, and location), used in census tables.
- *Address register* – records all addresses, by location, and identifies whether the address is used for accommodation, commercial (business) or office purposes. The address register may also include links to the building (such as an apartment building) and/or the dwelling.

2-61 There is no internationally agreed formal definition of an ‘address register’, and, indeed, the relationship between address and building or dwelling

registers (discussed further in Chapter 3) is often country specific. Some countries may have the address as an attribute in building and dwelling registers; other countries may have the building and/or dwelling as an attribute of an address in their address register. Some countries may treat each of them separately.

2-62 But whatever the national circumstances, the key condition for the census is that the NSO must have a list of places where people might live and be able to assign a location (ideally geo-coded) *and* link the place to the population register.

2-63 Without these base registers, it is very difficult to conduct a fully register-based census. However, even without a population register, it may still be possible to conduct a combined census, using data from other registers and information collected in a field enumeration as some of the examples in Chapter 4 will show.

2-64 While the final composition of these base registers, and their creation methods, will be country specific, there are some key pre-requisite requirements for their use for the purposes of the census:

- **Full coverage.** It is critical that the base registers meet coverage requirements. Examples of common coverage issues that would need to be resolved are the inclusion of people who are national citizens living in another country, and the exclusion of (a) people who usually live in the country but who have not yet formalized their residency status, and (b) children and young people. Similarly, where it is planned to create the base register from multiple sources, there must be no duplicates or major gaps in coverage.
- **Appropriate updating methods.** All base registers must accurately represent the situation at the census reference date, and not only at the date of registration. This can be a challenge for register sources which are not updated on a continuous basis, or where the updated reference date is not the same as the census reference date. For example, education registers are usually only updated at the start or end of the academic year. Where an NSO plans to use a previous census as the base for the statistical register, then it is necessary to determine how the register will be updated for all subsequent events.

- **Metadata.** As noted at section 2.2.6 in order for the NSO to properly understand the data, it is critical to have metadata that describes the sources and relevant variables.

2-65 Again, it should be emphasised that the national context is very important. It should never be assumed that the data collection methodology, including the use of administrative sources adopted in another country, can automatically be applied, as national circumstances, including the content and structure of administrative systems, are likely to be different.

2.3.7 Identifiers

2-66 As noted at section 2.2.7 above, in order to facilitate the linking of administrative data from different sources that relate to the same unit, it is preferable (if not a necessity) to have unique ID numbers at the unit record level common across all respective registers.

2-67 Such a unified identification system across different administrative data sources particularly facilitates register-based or combined censuses. The use of ID numbers to link data from different registers is discussed in more detail in Chapter 4. However, for countries where unique ID numbers for persons do not exist, the ability to link data efficiently and accurately is a particular challenge (noted again at section 2.4.6 below), and the NSO will need to develop appropriate alternative statistical methods (as discussed in Chapter 4).

2.4 Difficulties and challenges – and solutions

2-68 Despite the undeniable advantages that moving to a register-based census offers (as discussed in Chapter 1) such a change in methodology also presents the NSO with a number of challenges and risks that need to be addressed. Some of these have already been alluded to, but this section discusses them in more detail and offers some solutions.

2.4.1 Greater dependency on public authorities

2-69 In moving to a register-based or combined census, the NSO becomes heavily dependent on those public authorities holding the relevant

administrative records as the prime source of information. NSOs should recognise that, for such authorities, the production of statistics is not a core activity to which they would normally give priority. This could cause delays in the delivery of the relevant administrative data and metadata to the NSO despite any legislation obligation put on them to do so in a punctual manner. For the NSO, any failure or shortcomings in the administrative registers will affect the quality (and particularly the timeliness) of the statistical outputs, for which it alone must take responsibility.

2-70 Moreover, the fact that for wholly register-based censuses the NSO has to rely exclusively on the information originally collected for administrative, non-statistical, purposes inevitably imposes some restrictions with respect to the variables that are available for analysis and may also undermine international comparability. However, solutions to these potential restrictions can be sought by using information collected from pre-existing national surveys (such as a Labour Force Survey) and weighting the results to population totals in order to derive census variables that are missing from their registers (as is done in the Netherlands, for example). However, in such cases NSOs should be aware that the level of geographic detail in any subsequent analyses of such variables will be limited by the sampling fraction of the survey.

2-71 The use of registers also imposes on the statistical agency a dependency on the authorities responsible for the structure and maintenance of the registers as well as on any changes in legislation and administrative policy and practices. It should always be remembered that public sector administrative sources are generally set up for the purposes of organising public services, collecting taxes or monitoring government policies. This means that they are susceptible to political changes. If a policy changes, administrative sources may be affected in terms of coverage, definitions, thresholds etc., or possibly even abolished completely. Changes to the computer systems used to store and process administrative data may also have an impact on the supply of data for statistical purposes.

2-72 Such changes may happen suddenly with little warning, particularly at high-risk times immediately after a change of government, a change of minister, or a change in legislation. But even private sector sources (if they are to be used) are not immune to this sort of changes, though in this case, change is more likely to be driven by changing market factors.

2-73 Reliance on a particular source will always, therefore, carry a certain degree of risk. These risks can be managed to some extent by legal or contractual provisions. As noted at section 2.2.5, the best way in practice to avoid (or at least minimise) such problems tends to be through regular contact with those responsible for the administrative source, to ensure they are aware of the statistical requirements, and to try to influence, and get early warning of, any possible changes. Ideally, where changes to the content and/or structure of those administrative registers being used for the census are being considered, the NSO should be consulted in advance. Where there is a strong reliance on a particular source, it is also worth preparing contingency plans setting out what could be done if that source became unavailable. It is clearly better to be proactive beforehand than have to react after the event.

2.4.2 Differences in concepts and definitions

2-74 Registers and other administrative data sources often adopt different concepts and definitions of population-related variables than those that internationally apply in censuses (such as the concept of ‘usual residence’ or ‘unemployment’ or the definition of a ‘household’). NSOs should be aware that such differences may exist and decide whether or not these differences are acceptable within the national context. What may, in one country, be considered an acceptable difference when assessing the balance between the continuity and coherence of the resulting statistics and the reduction in field costs, may be considered unacceptable to users elsewhere.

2-75 One common example is related to the concept of unemployment. The standard statistical definition of ‘unemployed’¹ is:

“The ‘unemployed’ comprise all persons above a specified age who during the reference period were: (a) ‘without work’, that is, were not in paid employment or self-employment; (b) ‘currently available for work’, that is, were available for paid employment or self-employment during the reference period; and (c) ‘seeking work’, that is, had taken specific steps in a specified recent period to seek paid employment or self-employment.”

2-76 However, definitions of ‘unemployment’ in administrative sources are more often based on the number of people claiming unemployment benefits, or those registered as looking for work. Some people who are out of work may not register as unemployed if they expect to find work quickly, and in some cultures there may be a social stigma attached to claiming unemployment benefits. On the other hand, some people claiming unemployment benefit may not be available for work or actively seeking work, so should not be counted as statistically unemployed.

2-77 The first step towards solving the problem is to try to understand the differences and quantify the impact. Some differences may have no real impact in practice, so could be safely ignored, others may be systematic, so could be resolved through adjustments to the data. It may be the case that appropriate definitions and concepts can be approximated by derivations from different sources or by editing information from newly acquired census sources. In some cases, it might even be possible for the NSO influence the administrative definition¹⁵.

2-78 However, such solutions may not be possible and the NSO should then weigh up the balance between the acceptability of the differences and the costs of continuing a field enumeration for selected variables as required.

2.4.3 Classification systems

2-79 As is the case for particular variables, the classification systems used within administrative sources may be different to those recommended for the census. But, even if they are the same, they may be applied differently depending on the primary purpose of the administrative source, perhaps focusing on specific attributes of the unit. For example, an administrative source concerned with licensing, health and safety or environmental protection may be more interested in those economic activities of a commercial business that are of most concern to that source rather than the main economic activity of a business that, generally, is required for census purposes.

2-80 In other cases, classifications in administrative sources may not be applied at the level of detail required for statistical purposes (such as the 3-digit level of the International Standard Classification of Occupations ²⁰), or the classification may simply not be a priority variable for the administrative source at all, resulting in quality deficiencies.

2-81 Where classification systems or versions are different, the usual solution is to construct conversion matrices to map the codes in the administrative classification to those in the statistical classification. Such mapping may be one to one, many to one, one to many, or many to many. In the latter two cases, some sort of probabilistic allocation may be required.

2-82 Where accuracy is required at the micro-data level, the conversion matrix approach has severe limitations. Various other methods may be possible depending on resources and data availability, but a useful first step is always to gain a detailed understanding of how the classification data are collected and processed by the administrative source, and the nature of the administrative functions for which they are used.

2-83 In some cases, other variables may be available within the administrative source could be used to improve the likelihood of selecting the correct statistical code. One such variable could be the text description from which the administrative code was derived. If this is available, it is potentially of more use to the statistician than the administrative code itself, because the statistician could apply manual or automatic techniques to derive the correct statistical code directly from the description. This method can be used in conjunction with the conversion matrix approach, such that text descriptions are only coded in cases where there is not a one-to-one correlation between administrative and statistical codes, though there is a risk of potential bias if the quality of coding is different between the administrative and statistical systems.

2-84 One approach that has been used successfully in several countries is to develop an automatic coding tool for use in both statistical and administrative systems. This ensures a high degree of consistency of coding, and strongly encourages (but does not necessarily enforce) the use of a common classification system. In addition to the use of common coding tools, the provision of coding expertise and training to administrative data suppliers can help to improve coding consistency.

2-85 At the same time, it is always helpful for the statistician to stress the advantages of using a common classification system. It also helps to give early notice of any revisions to the classification system, and to provide as much help as possible to administrative data suppliers during the implementation of the resulting changes.

2.4.4 Inconsistency between sources

2-86 To compound the challenges arising from any differences in definitions and classifications between those adopted by the NSO for census purposes and those that are generic to particular administrative data sources, there may also be problems where there are inconsistencies between the registers themselves. Data from one source, for example, may appear to contradict those from another due either to different definitions or classifications (as previously discussed), or to differences in timing, or simply to an error in one source. The data on employment as recorded in a social security register, as one example, may not use the same business code as the taxation and business registers, and therefore extra processing is needed to link individuals to the company where they are employed. As another example, the linking of business enterprises to the building where they are based is not always straightforward since the company address data may not necessarily be fully accurate, or they may differ from the information in the building register.

2-87 To resolve such conflicts it is necessary to establish priority rules, by deciding which source is most reliable for a particular variable¹⁵. Once a priority order of sources has been determined for the variable, it should then be possible to ensure that data from a high priority source are not overwritten by a lower priority source. This process is made much easier if source codes are stored alongside variables for which several sources are available. The use and storage of date of the most recent update of a variable can also be helpful, as even when one source is thought to be more reliable than another, data from that source that are ten years old may be regarded as being of poorer quality than data for the most recent period from the (otherwise) less reliable source. A simpler method - that may be appropriate in some cases - is to load data in reverse priority order, allowing data of higher quality to automatically overwrite those of lower quality.

2-88 In most cases, there will be several variables of interest, and it is likely that the priorities will differ from variable to variable. For example, a register of employed persons is likely to give reasonable estimates of lawful employment, as that variable is closely related to the core function of the source. It may not, however, be so good for determining the economic activity of the employer, as this may be only of secondary importance to the register holder. Thus, if multiple sources are used to derive employment data, it would be necessary to consider the relative quality of each variable in each source in order to derive the optimal statistical data set.

2-89 The more data sources that are used, the more complex this comparison process becomes, but having multiple sources often helps to validate data quality. In some cases, certain sources may not be used directly for statistical production, but purely for benchmarking purposes as part of a quality assurance process. The resulting knowledge about the quality of various sources can also be fed back (usually at aggregate rather than unit level, to protect statistical confidentiality) to the source, and can provide a basis for discussions about improving the quality of that source.

2-90 A more detailed discussion on assessing the quality of administrative data sources is set out in Chapter 5.

2.4.5 Timeliness and reference periods

2-91 As has been made persistently clear in this handbook, public authorities responsible for maintaining administrative registers do not hold the data for statistical purposes, and, as a result, will have other priorities that could cause delays in the delivery of the relevant administrative data and metadata to the NSO. This can cause issues for the NSO regarding the timeliness of their register-based statistics, particularly where the timeliness of the delivery of data from different sources varies considerably.

2-92 There are three separate issues relating to timeliness that create challenges to the use of administrative data for statistical purposes. These refer to the fact that the administrative data:

- may not be available in time to meet data processing deadlines;
- from different sources may relate to different periods that do not coincide with the census reference date; and
- may be measured over a period, while the NSO's requirement is usually (at least for census purposes) for a specific point in time.

2-93 Considering the first issue, there will generally be some sort of lag between any event happening in the real world, and it being recorded by an administrative source before the data are made available to the national statistical organisation. This may be compounded by a further time lag caused by delays that might be experienced in the delivery of data by the register holder to the NSO referred to at section 2.4.1 above.

2-94 Lags relating to births, deaths and changes to place of residence are a major source of coverage errors. If these lags are measured allowance can be made for them in any census data derived from registers.

2-95 By analysing lags in this way, it is possible to produce summary statistics to estimate their impact. For example, it may be determined that two-thirds of births are registered within 40 days of their occurrence or that the median delay between occurrence and registration is 25 days. This sort of analysis is clearly important to help the NSO understand the nature and impact of the lags in the sources used to compile the census data. It also gives information that can be used to inform adjustments to improve the quality of the statistical outputs.

2-96 A further problem that NSOs encounter when moving towards a combined or register-based census is that different sources of administrative data often have varying reference dates. Sometimes a source gives the option of distinguishing clearly between reference dates and dates of events, but this good practice does not always apply. If these problems cannot be resolved sufficiently well, the risk is that not all sources will be harmonised to the same reference date and the key features of simultaneity will be lost (see section 2.5.2 below for a further discussion).

2-97 Then the question 'What is an acceptable difference in reference dates?' arises. However, the answer to this is dependent on the variable concerned. Some key variables (such as sex, age and address) are relatively stable over time such that a small difference in reference date is normally not a problem. But large differences in reference dates are always unwanted. However, it should perhaps be remembered that even in a traditional census, not all information may in practice refer to the single reference date of the census. If responses are received weeks after census day, recall effects may play a role and respondents may not always give answers specific to the census reference date.

2-98 The third issue concerns the difference between data referring to a specific point of time and data relating to an extended period. For example, there will usually be a census requirement for employment data to refer specific reference date or a short reference period immediately prior to census day (usually a week), whereas administrative data may only provide monthly or annual averages.

2-99 As in previous examples, the first step is to analyse the impact of the difference, and determine whether or not it is significant enough to require

further action. One possible solution is a model-based mathematical adjustment – for example if the census reference date is near the start of the month, then a model that takes into account the average figure for the previous period may be appropriate. An alternative approach may be to use the results of a relatively small survey to adjust the administrative data.

2.4.6 Lack of a unified identification system

2-100 All statistical units in a register-based census should be linked to one another by means of identification systems: there should be the capacity to link persons to household-dwelling units and to the dwellings and buildings in which they live, and for employed persons to be linked to their employers. Employers and buildings also need to be linked in order to determine workplace.

2-101 As has already been noted (at sections 2.2.7 and 2.3.6) one major factor that facilitates the linkage of disparate administrative data records is the application of unified identification systems across different sources. The data linkage must occur at the individual level. While it is theoretically possible to link records on the basis of the name and other unique details of the individuals (such as age, sex and address), the existence of a unique identification number for each individual, household and dwelling allows a much more effective and reliable linkage of records from different registers. Indeed, in the absence of such a unified system and an established central population register it is extremely difficult and laborious, if not impossible in some countries, to link data from different registers which is a pre-requisite for register-based statistical production. Similarly, all units should be geographically located by using local area codes or map coordinates.

2-102 There are, however, alternative approaches to linking data sources that can be adopted where unique ID number do not exist. These are discussed in Chapter 4.

2.4.7 Missing data and the under-coverage of some populations

2-103 The problem of missing data is not unique to administrative sources. It can also be due to full or partial non-response to statistical surveys, or even to the removal of data values during the editing process¹⁵. However, with

administrative sources, the issues can sometimes be different, particularly as the problem of missing data can often be more systematic.

2-104 The main reasons for this are that a particular variable may not be collected for the entire target population but only be collected for certain categories of units for which there is a specific administrative requirement, or may not be collected at all by the administrative source. The variable may also simply be a low priority for administrative purposes, so the owners of that source do not see missing data as a problem. Moreover, even for those countries that use sample surveys to collect data on information not available in administrative sources, it is sometimes difficult, or even impossible, to produce accurate outputs for small areas or specific sub-groups of the population because of the size of the sample population.

2-105 Some of the standard solutions for dealing with non-response in statistical surveys can also be used to solve the problem of missing data in administrative sources. Various imputation methods, such as deductive, 'hot-deck' or 'cold-deck' imputation are often suitable where the problem only affects some of the units. In cases where most, or all, of the units are affected, a modelling approach may be more appropriate. (Data quality issues are discussed in more detail in Chapter 5.)

4.4.8 Maintaining IT infrastructure

2-106 In countries that conduct censuses with long time intervals between them, it may be difficult to retain staff at the NSO with the necessary experience and expertise to keep the knowledge and IT infrastructure up-to-date during the inter-censal period. However, when yearly census updates are introduced (made possible with the use of regularly updated registers) this difficulty is minimised.

2-107 It is generally recognised that the numbers of IT staff (and, indeed, other staff) required to manage a register-based census on a day-to-day basis can be significantly fewer than for a traditional census. However, in reducing the staff numbers required to a minimum, the NSO may lose much long-acquired experience in, and knowledge of, the census that will need to be replaced through appropriate training - as noted at paragraph 2-35 above. Remember, however, that wherever a significant element of a traditional field operation is retained (in a combined census approach) the level of previously acquired IT

skills and experience necessary for data collection and processing will need to be retained.

2.4.9 Privacy and security concerns

2-108 As noted briefly at section 2.2.4 the existence of more and more administrative registers in society may trigger discussions on privacy issues. If the public attitude should become negative, politicians may become reluctant to establish new registers or upgrade existing ones. It is extremely important, therefore, that the general public appreciates and understands the benefits of using register sources for statistical purposes. It is vital that the national register legislation is up-to-date and that the work of the register authorities is open and transparent.

2-109 Although there is often a general unease amongst the public about data sharing, there are also contradictory pressures to improve the efficiency of government, particularly if this results in lower taxes or more funding for key services such as health or education in voter-popular areas. Political slogans such as 'joined-up government' are often appealing to the public, and can help to counter fears of loss of privacy. Thus, the extent to which improvements to public image can be seen as an advantage of using administrative sources depends heavily on how that use is presented to and perceived by the public. As Wallgren and Wallgren have noted¹⁹, if the NSO has a good reputation for trustworthiness, it will be easier to gain public acceptance of the use of administrative data for statistical production. However, one indiscretion in the protection of its integrity can easily destroy any such reputation.

2-110 People can perhaps be educated to appreciate that administrative authorities are already collecting, as a matter of course, the same data that the NSO could use for statistical purposes – for example: the tax authorities hold information on everybody's employers and earned income; work and employment authorities hold information on all the working hours of employees; labour market authorities hold information on length of unemployment; and pension institutes hold information on old age and other pensioners. In this situation it is possible to encourage the public to accept the argument that using such administrative data for statistical purposes has the benefits of cost efficiency, reduced public burden, and improved data security (because of the reduction in the number of persons handling the data, both within the NSO and externally and that all such data is now in an electronic format).

2-111 However, it is important that the statistical agency always remains on its guard in this respect. It is very easy to lose the confidence of the general public, but a major effort to regain it. NSOs should therefore be prepared to respond to specific concerns about privacy, confidentiality and security issues in conducting a register-based or combined census. Where a combined census is being contemplated a discussion can be expected about which variables are to be included on the census questionnaires and which variables are to be derived from administrative data sources.

2-112 Clear and well publicised protocols in the use of administrative registers for statistical purposes are vital for this support. Strong frameworks and clear provisions for protecting confidentiality and assuring privacy of individuals should be reinforced by robust practices and infrastructure. This should include very explicit provision on access, as well as clear oversight, redress and correction processes if things go wrong¹⁸.

2.4.10 Resistance to change

2-113 One of the main barriers to the more effective use of administrative sources in official statistics and the census in particular, and perhaps one of the least recognised, can come from within the NSO itself. Statisticians brought up on traditional data collection methods may resist the use of administrative sources because they do not trust data that they have not collected themselves. They often focus on the negative quality aspects of administrative data, and they have an over-optimistic view of the quality of data collected in the field¹⁵. They may also fear the threat of their own redundancy.

2-114 The solution is clearly through better education of statisticians regarding the possibilities offered by administrative sources, encouraging them to take a wider view of all the dimensions of quality, and focus on the impact on data suppliers and users. In this context it is important to determine the real relative quality of traditionally collected and administrative data. For example, it is often assumed that data from administrative sources do not meet the requirements of statistical definitions, whereas those from specifically design censuses do. Indeed, sections 2.4.2 and 2.4.3 suggest that this is often the case. However, there may not be any real difference in practice, particularly if respondents to self-completion census questionnaires do not read the often-lengthy notes about how a particular variable should be defined for census purposes.

2-115 A further way to help break down the barriers of internal resistance is to show that cost savings from using administrative data do not necessarily mean staff reductions. The resources saved can, at least partly, be used to improve quality or increase the range or frequency of outputs.

2.4.11 Diminishing interest

2-116 In countries where censuses are carried out using questionnaires, not only the users but the general public itself will be interested in knowing the results. However, in register-based census countries, where people no longer complete census forms, there is often a decline in public interest in (or even awareness of) the census and its results. As a consequence, national interest in, and perceived importance of, the census is greatly reduced.

2-117 Users will still have an interest in the statistical outputs, though evidence from register-based census countries suggests that their interest in the choice of original sources and the methodology used to produce the census data diminishes over time. This is also due to the fact that in those countries other outputs are often available earlier than the official census results. Such diminishing interest is seen by some as a harbinger of the end of the census as the principle source of socio-demographic information in the long term.

2.5 Impact on the essential features of the census

2-118 In moving towards a register-based (or combined) census methodology, NSOs will want to assess the impact that reliance on administrative data sources will have on the essential features of the census discussed at section 1.4.3.

2.5.1 Individual enumeration

2-119 The principle of *individual enumeration* is a fundamental feature for any population and housing census of population. Traditionally, this has been ensured by providing questionnaires that ask questions of each individual within a household. In the case of register-based or combined censuses a different approach is adopted for which it is important that each census unit (that is, the individual person, household, or dwelling) has a special, uniquely identified record in the registers used.

2-120 If a single identifier for a particular unit does not exist across a range of registers, it is necessary to create a new statistical identifier (based on a group of identifying variables) to link the variables held in the respective registers, and to carefully check its quality (for errors and uniqueness).

2-121 Sometimes it is necessary to derive census variables by combining information from several administrative registers using special algorithms. This is possible if the units in all these registers are uniquely identified by the same identifier. In this case the variable created in such a way should be uniquely identified as well and saved in a statistical register.

2-122 Subject to this ability to link data and caveats noted in this Chapter , the feature of individual enumeration is satisfied in the register-based census approach as separate information is collected regarding the characteristics of each enumeration unit.

2.5.2 Simultaneity

2-123 Traditionally, to ensure *simultaneity*, the enumeration is carried out over a very short time frame, ideally during one day only, but generally this is not practicable. Though most modern-day enumerations are conducted over a two-three-week period, all the data collected should refer to the same specified reference period. This essential feature should be respected also in the case of a register-based or combined census. However, the concept of simultaneity may require careful consideration, since even if data is extracted from the various registers at the same time, register update cycles vary and the data may not always relate to the same reference date.

2-124 If the administrative registers in use are regularly updated, then it is necessary to fix the census period and to take the data from all registers with reference to this period. Sometimes the registers are updated regularly at some specific date such as the beginning of a year, and then it is possible to use this date as the census period, and the simultaneity of the census is guaranteed. In the case of a combined census, it is important that the census reference period to which the questionnaires refer and the reference period of the information taken from the registers are the same or as close to each other as possible.

2-125 When several administrative registers are used in the census, it is important that all data taken from them have the same reference period. Usually, census variables that are derived via special algorithms take some time

to calculate; hence those census variables are only ready for publishing sometime after the census period.

2-126 For some specific variables in combined or register-based censuses different reference periods may be defined for the particular administrative purpose of the register. For example, demographic data can normally be taken from population registers at the beginning of the year. However, labour force data might be more relevant some time before the end of the year - as fewer people tend to be in employment around Christmas and New Year's Eve.

2-127 For some administrative registers, it may not be possible to have Census Day as a reference day. For example, education registers often have relevant education data (referenced to a day early in the academic year) that may differ from a chosen Census Day. In such cases, the NSO may wish to take education data with a reference day as close as possible to Census Day as a compromise.

2.5.3 Universality

2-128 To ensure the universality of the traditional census the questionnaires used in the enumeration process are the same for all households and individual persons.

2-129 If the administrative registers used in the census are common for the whole country and all population groups, or if the enumeration is taken from a population register in which the fields for attributes are populated from subsidiary registers relating to specific topics, the criterion of universality may be regarded as being met. However, if there are different administrative data sources in different areas or for different population groups (such as an urban population register and a rural population register, or if different registers are maintained in different cities), then it is necessary to analyse the possible discrepancies between the different administrative sources and find a way to define common census variables using these different administrative sources. In this case plausible results can be derived from these newly defined variables (via an appropriate algorithm in a statistical register).

2.5.4 Defined periodicity

2-130 *Periodicity* is clearly satisfied, since through a register-based approach extraction of the data and hence publication of the statistical results can be at

any desired frequencies, including the traditional five- or ten-year cycle. Indeed, a key advantage of a register-based census is the opportunity to conduct the census (and hence release the results) more often than the traditional decennial cycle, as register data are permanently available and more regularly updated. Then the periodicity between censuses can be ten, five, two or even single years.

2-131 It is also possible to produce some census updates with a reduced list of variables (but long enough to meet users' requirements), which could save resources. From this it follows also that in countries where a regular decennial census will be continued (when using a combined methodology for example), a shortened list of variables derived from administrative data sources can be updated more often.

2.5.5 Small area data

2-132 Providing a rich wealth of information for small geographic areas and small population sub-groups (generically referred to here as 'small area data') is a key objective for a census of any kind, as there is generally no other single source of comparable data available to the NSO. The demand for geo-referenced data on the population as well as buildings and dwellings has increased significantly in recent years to the extent that has resulted in the introduction in the UNECE's recommendations for the 2020 round of censuses of a new 'population grid' non-core topic³.

2-133 In the context of register-based or combined censuses, small area census data can be derived from administrative data providing the coverage is high, preferably covering the whole population. If there are some small areas that are poorly covered, resulting in the administrative data lacking some information (and thus showing poor universality), it will be necessary to seek to improve the administrative dataset before it can be used as a source for the census. Such poor coverage is likely to be a problem for the every-day usage of the administrative data, and so it should be improved anyway. Improving the statistical register can sometimes be done by adding information from another source, providing that linkage through common IDs is possible.

2-134 Sometimes there might be special administrative data for some small hard-to-enumerate sub-population groups (such as refugees, nomads or the homeless). In that case it will be necessary to combine these different administrative sources. If the result is satisfactory, this combination is useable.

2-135 In the case of a combined census, it is also possible to supplement the lack of information in administrative sources with a survey that, if necessary, may use different data collection methodologies, such as doorstep or telephone interviews, or self-completion with paper or online questionnaires, to suit different areas. However, in this case where a sample survey is used rather than a full-field enumeration, problems can still arise with the coverage of small area data. In such circumstances, users' needs regarding the required level of detail of the census outputs should be taken into account before any decision is made on the sample size of the census survey.

3 Types of administrative sources

3.1 Introduction

3-1 Integrating data from various sources in a population and housing census is becoming the increasingly adopted approach among NSOs worldwide. In countries with administrative registers that use unique identifiers for unit records, it is important to distinguish between the key (or base) registers and supplementary registers (which are sometimes referred to as specialised or subject-matter registers).

3-2 Base registers are those that hold the basic information relating to the stock of the entire set of 'population' units that are being counted in the census – typically persons, dwellings and buildings. Each unit is usually identified by a unique number such as a PIN, real property number or a dwelling number. The specialised registers hold the record for each unit, often identified by the same unique number as in the base register, with information necessary for the specific administrative function of the data holders, and from which one or more of the characteristic variables that may be required for the census (such as educational attainment, occupation, disability or income) can be derived. Such specialised registers may not necessarily cover the whole of a census's target population.

3-3 This chapter focuses on the content, structure and requirements of the base registers and use of the other specialised registers for the purpose of deriving census attributes. It concludes with a summary of the design of personal identification numbers used in the linkage of information contained in such registers.

3-4 However, it should be noted that for the purposes of the census it may be necessary to integrate data not only from other administrative registers but also (as in the case of a combined census methodology) from other data sources, such as a field enumeration, on-going sample surveys (such as a labour force survey), ad hoc surveys or a longitudinal study of a census-based sample. The integrated dataset from such sources assumes the form of a statistical dataset by which is meant a database that can be used in the further process of compiling and maintaining data for the purposes of a census (or for any other statistical purposes).

3.2 The base registers

3-5 When transferring from a traditional census to a combined or register-based census methodology, the establishment of base registers, in particular a population register and building/dwelling registers, is of fundamental importance.

3-6 As noted above base registers are those that hold the basic information relating to the stock of the entire set of ‘population’ units that are being counted in the census – typically persons, dwellings and buildings. Base registers are created and then linked at the unit record level to the information taken from specialised registers (sometimes referred to as subject-matter registers) from which one or more of the characteristics that may be required for the census can be derived) in order to create the requisite census attribute variables of the population and dwellings from which the census statistics are then produced.

3-7 This section of the handbook discusses the base registers that are most commonly used for the purpose of the census, namely:

- population registers
- dwelling or building registers
- enterprise or business registers
- address registers

3.2.1 Population registers

3.2.1.1 The concept of a population register

3-8 The term ‘population register’ was originally defined by the UN as “*an individualized data system, that is, a mechanism of continuous recording, and of coordinated linkage, of selected information about each member of the resident population of a country in such a way to provide the possibility of determining up-to-date information concerning the size and characteristics of that population at selected time intervals*”²¹. However, such a definition is not necessarily universally adopted. For example, the Organisation for Economic Co-operation and Development describe population registers as: “*Accounts of residents within a country. They are typically maintained via the legal requirement that both nationals and foreigners residing in the country must register with the local authorities. Aggregation of these local accounts results in*

a record of the population and population movement at the national level"²⁹, while the World Bank prefers the definition: "A database of every individual that has the right to reside in the country, including citizens and non-citizens, children, and adults. Population registers typically contain demographic data and life-event information that is the basis of or exchanged with other identification systems and databases such as national ID systems, civil registers, and others"³⁵. A further variant is defined by the Organisation for Security and Cooperation in Europe as: "the system provided for in a consistent legal framework setting out terms and conditions to continuously register eligible persons within a specific area of a public authority with the purpose of establishing their identity, civil status (including vital life events) and place of residence, and to provide them with proof thereof on the basis of documental evidence."

3-9 Put more succinctly, and for the purposes of this handbook, a population register is taken to mean an *administrative register* primarily used in an administrative information system by public agencies for their own specific purposes (such as establishing personal identification, voting, education and military service, social insurance and welfare, and for police and court reference. Register information is also utilized for issuing documents needed for the admission of children to nurseries, kindergartens and schools and the assignment of residents to health clinics) *that provides a frame (or list) of the resident population (however defined) in a given country or area*. Additionally, it often provides some demographic characteristics of individuals such as date of birth and sex (see paragraph 3-14 below).

3-10 It should be noted that the concept of the term 'resident population' referred to in the previous paragraph varies from one country to another, depending on the purported use of the population register. The register normally includes citizens and non-citizens who are legal residents of the country, and it is not uncommon to include citizens living abroad. Individuals residing temporarily in a country are, however, not included among the 'resident population'. For people to qualify for inclusion, a minimum period of stay within the country is commonly specified.

3-11 Thus, a population register is the product of a continuous process, in which notifications of certain events, which may have been recorded originally in different administrative systems, are regularly recorded in the register. The method and sources of updating should cover all changes so that the characteristics of individuals in the register remain current. Because of the

nature of a population register, its organization, as well as its operation, should have a legal basis.

3.2.1.2 Purpose and content of a population register

3-12 In their administrative form, population registers are usually held by national or local government departments or authorities. They provide a systematic list of persons - usually residents of the country, but in some cases only national citizens (see sub-section 3.2.1.3 below). They are typically maintained to fulfil any legal and administrative requirements that both national citizens and foreigners residing in the country should register with local and national authorities. It is important, therefore, that such registers are continuously updated. Where local authorities administer the population register, then an aggregation of the local registers results in a record of the population at the national level (assuming that such registers are harmonised to avoid any duplication of unit records).

3-13 The primary function of a population register is to provide reliable information on who is, or should be, in the country for the administrative purposes of government, particularly for programme planning, budgeting, civil and vital registration, and taxation. They are typically maintained to fulfil any legal requirement that both nationals and foreigners residing in the country should register with the local authorities. It is important therefore that such registers are continuously updated. Aggregation of these local accounts results in a record of the stock of the population at a point in time at the national and local level. Such registers are also useful in other administrative areas, such as:

- establishing personal identification;
- the administration of pensions, health, educational and other services;
- establishing ownership of a residence or the right to work; and/or
- creating lists of electors for the purpose of the management of voting at elections or in a referendum.

3-14 The basic individual characteristics that may be included in a population register are determined by the particular political and/or administrative use to which the register is put. The most commonly used variables are:

- name,
- sex,

- date and place (or country) of birth,
- marital status,
- personal identification number,
- citizenship, and
- address of usual residence and date of arrival at, or departure from, that address.

3-15 The registers on some countries may also include details of any spouse, any children and in many cases the names (or IDs) of parents if these are required for the purposes of identification. The Nordic countries, for example, include the PINs of the parents as these are important for many administrative functions. They are also useful for making family statistics and for scientific research – for example on hereditary diseases.

3-16 Other countries, however, may prefer to include as little information as possible in the population register in order to avoid recording characteristics that are liable to change, thereby reducing the amount of effort required to keep the register up-to-date. In such cases there would need to be a much greater reliance on, and linkage to, specialised registers to provide the requisite census characteristics.

3.2.1.3 A centralised population register

3-17 A centralized population register can be a network of local registers that are linked in a coordinated way. Although the national population register may be a virtual entity based on the linkage of population registers established at the local level (decentralized system), the overall geographical coverage must be of the entire territory of the country. If this condition is not met, the national population register will not be an appropriate system for the production of statistical data for the country.

3-18 Likewise for the territory, the entire population must be included in the population register(s), either central or local. On the other hand, a population register may induce over-coverage errors if data are not properly filtered during the data compilation process. For instance, a decentralized system based on local registers may incur a higher risk of duplication of records of individuals when summing up data at the national level.

3-19 The historical development of a central population register in different countries has depended heavily on national circumstances. In many countries the creation of such registers was to satisfy the several requirements and administrative responsibilities of local and national authorities.

3-20 The initial step towards establishment of a population register is to build a stock of the existing resident population. There are two ways of doing this: (a) by using existing databases to consolidate a list of residents; or (b) registering eligible residents through a one-time mass registration. Some countries (Norway is a case in point) created a purpose-built register from the information collected in its previous population census, including all persons that had been enumerated. However, the use for any administrative purposes of register created in such a way would require particular national legislation and would, it should be noted, now be contrary to current internationally accepted good statistical practice. Indeed, it is especially important for the NSO to maintain a high level of data protection and to ensure that the information collected for the purposes of the census in particular cannot (and will not) be used by the NSO for anything other than statistical analysis.

3-21 However created, when centralised, a population register is now used in many countries as a base register for statistical purposes and provides a key component of any register-based statistical system. It can provide the basis for determining the socio-demographic profile of the population at any point in time as well as measuring the components of population change – births, deaths and migration (both internal and international) - in order to produce national and local population estimates and projections. To fulfil these functions, and especially for using the register as the base for a census, it is necessary for the NSO to ensure that the register is regularly updated (see section 3.2.1.5 below).

3-22 A discussion on the creation of a statistical population register within the NSO is given in Chapter 4. However, it can be noted here that any statistics on population and vital events derived for the population register for the purposes of the census should refer to the usually resident population. While for administrative purposes it may be legitimate to include in the register persons who are not usual residents of the country (such as citizens living abroad and temporary residents), for census purposes care should be taken to identify the correct target population, especially if the statistics are to be used for international comparisons.

3.2.1.4 Who is to be included in a population register: the problem of establishing the place of usual residence

3-23 As noted above, a population register will usually be required to include all national citizens and other residents who live in the country (or in specified local area). Ideally, coverage should be comprehensive but, inevitably, there will be a degree of under-coverage arising from omissions due, in the main, to the difficulty in recording international migrants (especially emigrants and undocumented immigrants) and the homeless, and of over-coverage arising from duplications where, for example, people have more than one residence.

3-24 For each person in the register a specific place of residence must be recorded – but the criteria for determining place of residence and the level of geographic detail (for administrative purposes) will vary from country to country. Even determining whether or not a person qualifies as a resident in the country may be problematic.

3-25 Within Europe, for example, countries adopt different concepts for defining the resident population for the purpose of the census, reflecting the administrative use to which the information is put. Some use the current internationally accepted statistical concept of *usual residence* based on an established period of actual stay or intended stay (usually 12 months) while others adopt the concept of *legal residence* where the legal population (sometimes referred to as the *de jure* population) is defined as those persons who are entitled to be settled in the same country at the census reference time, either by holding the national citizenship or by some other authorization issued by national authorities. Those countries where registers are used for the census may, instead, choose to adopt a *registered residence* concept in which the population base comprises those persons who are listed in one or more registers as living in the country at the time of the census whether or not they are actually in the country.

3-26 Citizens who are living abroad will include some persons who should be treated as only temporarily absent if their stay away from home is shorter than the defined period that would qualify them as an emigrant. Such persons (such as students who travel overseas for further education or people who work overseas for less than the minimum duration required to be a resident) should perhaps rightly continue to be included in the register. But those who have left the country with the intention of staying away for a longer period should, in principle, be removed from the register (and thus not recorded in the census).

However, in practice this often does not happen since there may be few incentives for emigrants to report their intentions to the relevant administrative authority. In addition, many emigrants may still retain active links to their home country (through, for example, home ownership, tax obligations or family commitments) that discourage them from considering themselves as being usually resident abroad).

3-27 Conversely, foreign citizens living in the country (and any national citizens returning from abroad) should be included in the register if their length of stay exceeds the qualification period for residency, but excluded otherwise. In most cases the coverage of foreign nationals in the population register is comprehensive since it is generally to their advantage that their residency in the country is officially established in order, for example, that they can obtain work permits or entitlement to benefits and access to health, education and other services. However, it should be noted that the recognition of the presence of refugees and undocumented migrants – even if permanent – presents particular problems for many countries. Moreover, as in the case with national citizens who emigrate, there is the possibility that some registered foreign citizens who subsequently return to their home country without informing the administrative authorities, will be erroneously retained in the register.

3-28 Clearly then, the NSO needs to fully understand the rules and practices that apply in the creation of the population register, and to determine the extent to which additional work will be needed to ensure that they meet the statistical requirements.

3.2.1.5 Population registers and civil registration and vital statistics systems

3-29 The foundation of a population register is based upon a civil register that maintains records of vital events such as live births, deaths, marriages, and divorces, and which can be used to derive statistics on demographic flows. While a civil register on its own is not a population register, it is a fundamental element a functioning population register in that as a minimum requirement all births and deaths should be registered. If vital events are not registered and the records are not reflected in the population register, then the ability to update the population registers in order to generate census statistics is limited¹⁸.

3-30 Consequently, while the primary purpose of a population register must be administrative, it is also a potentially excellent source of data for generating vital

statistics and key demographic indicators. Furthermore, some population registers need to increase coverage and data quality in order to be a suitable data source. Generally, where population registers are founded on well-functioning civil registration systems, it becomes possible to expand the accuracy, reliability, granularity and scope of the statistics which can be derived from them.

3-31 A significant advantage of using population registers to compute vital statistics is the possibility of directly calculating specific demographic rates with potentially no numerator-denominator bias. For instance, it could be possible to compute specific fertility rates for employed and/or immigrant women, parity progression ratios, life expectancy by educational attainment, indicators on mixed marriages by ethnic group/foreign background, divorce rates by socioeconomic class of the spouses, etc.

3-32 This requires full matching between civil registration and population register data, with the civil registration data serving as a fundamental data source for the population register. Additions to the register through births and immigrants and removals through deaths and emigrants are thus fundamental movements that need to be recorded; completeness of the register is key to the production of good quality vital and demographic statistics. Moreover, the same level of detail of information should be contained in both sources, meaning that the certificate of the event (birth, etc.) must contain the same classification of information as that available in the population register. In general, the use of the population register provides a broader opportunity to correctly identify the population segment at risk of an event. The timeliness of the updating of the population register and the accuracy of the information recorded therein are therefore factors critical to the quality of the statistics to be computed³¹.

3.2.1.6 The role of population registers in measuring migration

3-33 In addition to recording the stock of population in any country, one of the major purposes of a population census is to provide statistics that give an insight into the size and distribution of the population and its changes. Moreover, an important feature of a census is that it can (and indeed must) provide statistics for small areas and regions.

3-34 Though not considered a vital event (as described in section 3.2.1.5 above), migration is also a critical data element in a population register whose inclusion enables the size of population to be accurately derived at any given

point in time and for given geographical units. (See the following section 3.2.1.6 for further discussion of the role of that the population register plays in measuring migration.) If the census is going to be based on registers (either wholly or partially), the major data source will be a population register that can provide reliable and continuous data on where people live at any point in time, preferably the address or dwelling number, or at least the locality, such as village or municipality. As noted at section 3.2.1.5 above this source would have to be kept up to date with data on births and deaths from the country's civil registration system (CRS). However, the third component of population change – migration (which in some countries may be as significant an element in population growth as births and deaths) - is usually not recorded by the CRS.

3-35 Migrations are generally more complicated to define and record than births and deaths. Most births and deaths occur in health institutions - depending on the local situation - and professional personnel are usually involved in the issuance of notifications or certificates of these events. Changes of address, on the other hand, often have to be registered by the migrants themselves, and thus involve subjective judgments and actions. Unfortunately, there are fewer incentives for registering migrations - in particular emigrations - than for births and deaths. There may even be disincentives for *not* registering a migration under particular circumstances (as noted above) - often driven by the fear of losing social and financial advantages that are associated by being registered in a country or a region within a country.

3-36 Another complication is that the *duration* of a move determines whether the change of residence may be classified as a migration or not. In principle, it is the duration of the *intention* of the move that is significant in this respect. According to the recently revised conceptual framework³⁰ an international migrant is defined as a person who has changed his or her country of residence (as defined by the UN *Principles and Recommendations for Population and Housing Censuses*¹) and has established a new residence in the country within a given year. For *internal* migrations the definition is also based on a change of place of usual residence using the same criteria as recommended by the UN. It should be noted that for both international and internal migration, countries may use different criteria depending on their national legislation and practice.

3-37 The most important source of information on internal migration is self-reported change of address. This is done in many countries, including Norway, Bahrain, Kyrgyzstan, Turkey and the Republic of Korea. Moves are usually reported to the same population registry that registers the vital events of births,

deaths and marriages. This may be a legal requirement, but there may also be incentives for doing this in order, for example, to be employed, or to access social services such as health and schooling, or to receive utilities such as water, gas and electricity. Migration numbers can be estimated from such notifications according to the definition of usual residence and the desired category of internal migration - between regions, counties, municipalities. Migration within the *same* administrative area may also be estimated if the information on address is sufficiently specific.

3-38 Possible sources of migration data include:

- immigration departments or agencies, including offices issuing visa and residence permits;
- passport control authorities at border posts at airports, seaports and elsewhere (see also section 3.4.6 below);
- cooperation between countries in exchanging information on immigrations and emigrations - examples of this is the cooperation between the five Nordic countries and between the three Baltic states.
- tax agencies, particularly for internal moves; and
- individual reporting of internal or external moves.

3-39 In principle, it is possible to estimate annual migration flows by comparing the addresses of everybody in the population register at the beginning and end of the year. However, it would usually be easier to use the notifications or records that are used to update the address data in the register as described above.

3.2.1.7 The role of population registers in identity management

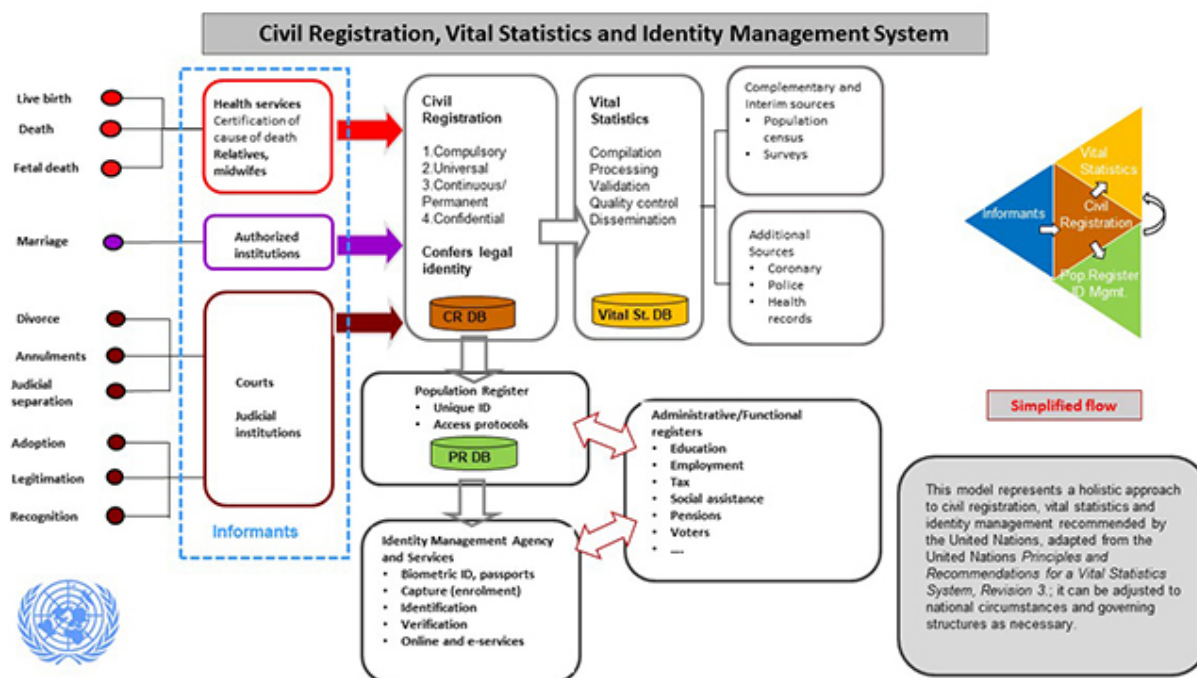
3-40 Legal identity is acknowledged to be catalytic for achieving at least ten of the Sustainable Development Goals (SDGs). Data generated from civil registration and population registers support the measurement of over 60 SDG indicators. As civil registration establishes the existence of a person under law, it has traditionally been the fundamental means of granting legal identity. Legal identity has a critical role to ensure the global community upholds its promise of leaving no one behind as espoused in the 2030 Agenda and should be inclusive of all people from birth to death²².

3-41 While there is no internationally agreed definition of **legal identity management**, the term refers to the issuance of a proof or legal identity to each

individual and the maintenance of systems for managing information and documents associated with such identity. It also comprises verification, registration, management, and conservation of personal data of citizens (as well as non-citizens) within the state territory with the goal of establishing a unique legal identity credentials.

3-41 The United Nations Legal Identity Agenda, launched in May 2019, consists of a holistic approach to civil registration, vital statistics and identity management and represents a model that can be used by all member States. The population register is a major component in that model (where it exists) and there should be full interoperability between these functions in a simultaneous manner, according to international standards and recommendations and in compliance with human rights of all people concerned, including the right to privacy. The essential premise is that the civil registration system is uniquely positioned to provide legal identity data to be entered into the population registers including additional layers of identity data as a consequence of registration of vital life events through lifetime.

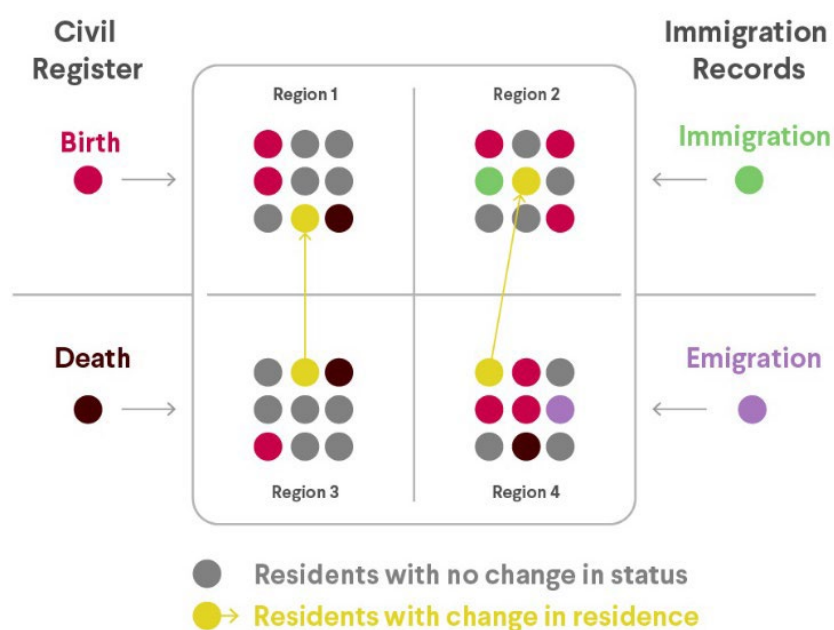
3-42 The UN recommends that all Member States should adopt and implement this agenda as a systematic and perpetual mechanism for ensuring legal identity for all³². The graph below represents the model of a holistic approach to civil registration, vital statistics and identity management.



3-43 Once established, legal identity information in the personal record in the register is subject of continuous updating of the facts of births, deaths, adoptions, legitimations, marriage, divorce, annulments and judicial separations, change of name or sex, and change of residence. An efficient connection with the civil registration authority is therefore a fundamental element for the proper functioning of the population register.

3.2.1.8 Updating and ensuring the quality of a population register

3-44 Once the stock of all currently living residents is established, it is continuously (in some cases in real time) updated, using a flow of information on vital events registered in the civil registration system, as well as records of migrations into or out of the country. A resident is entered in a population register when her/his birth or in-migration is registered and is retired from the register (that is, marked as dead or moved out of the country) when his/her death or out-migration is registered (as illustrated in the Figure below). Note that no record should be entirely deleted from the population register; records of the deceased and out-migrants should, however, be appropriately marked²⁷.



3-45 In addition to the information from the civil registration of births and deaths of residents, characteristics of these events, such as sex, date and place of occurrence, are also recorded in the civil registration and vital statistics (CRVS) system and transferred to the population register. New entrants, other than new-borns, such as those left out during the establishment of the register and new in-migrants, are registered in the population register, based on the submission of documents as specified in the relevant law and issued by designated authorities.

3-46 While the establishment and maintenance of population registers rely on information collected by administrative databases (mainly the civil register and immigration database/register of residents), the architecture of the registers should not be seen as a stand-alone database that permanently copies information from other registers. Rather, they can constitute one coherent database or a system of interconnected and interoperable databases, often at different locations and operated by different authorities. For example, in the Nordic countries, the civil register and population register are one entity and any changes in civil status are directly entered in the population register.

3-47 As noted at section 3.2.1.5 a population register cannot be described as such without it being linked to the civil registration of vital events, which constitute information fundamental to its continuous updating, together with

changes in physical address. It should be noted that the physical address should be accompanied by geographic location particulars, such as province, district, name of the village or town (or even ward within the town). However, it is recognised that many low- and middle-income countries may not have any physical address system, particularly in rural areas, in which case just the location particulars mentioned above are required. The geographic location codes should be part of the database so they can be used for compilation of statistics at various of geographic levels (see also section 3.2.4 on Address registers).

3-48 Also, as noted at section 3.1.2.6, a population register maintains information about an individual's current residential address and records any changes in address resulting from entry into a country or change of residence within a country. While maintaining such updates of internal movements remains a vital aspect of maintaining a population register, a change of residence or address within a country is usually voluntarily reported which makes it difficult to maintain accurate records and data on internal movement. In a few countries, however, it is a legal requirement to provide formal notification of one's departure and to register at one's new place of residence. There are also countries where addresses in the population registers are updated using the information in address databases maintained by separate agencies designated by law.

3-49 In general, the quality of central population registers is constantly improved as the result of feedback from users who have access to the database on a continuous basis. It is clearly in the interest of both local and national administrations to ensure that the status of the individuals recorded corresponds to reality – as well as to the law.

3-50 But no matter how much care is taken and how many checks are made, errors occur, sometimes on a recurrent basis. These can be due to incorrect information either reported by the individuals or recorded by administrations concerned, or resulting from a *failure* to report, as in the case of emigrations. Although in most countries the data recorded in the administrative register are not the responsibility of the NSO, statisticians need to thoroughly examine the data provided by the register owners and inform them of the extent of any deficiencies or inconsistencies that they discover.

3-51 In reporting any such inconsistencies, it should be stressed that the guiding rule is that the NSO should only advise the administrative agency about

systematic issues and (in order to be seen to be protecting the confidentiality of census information) should never report instances of errors or omissions relating to individual unit records.

3-52 All or most of the data items held in the administrative register are recorded on the basis of reliable documentary evidence. For example, for those related to vital events as noted in section 3.2.1.5 above:

- births (including, name, sex, date and place of birth, and details of parents);
- marriages (including names, date and place of occurrence and previous marital status);
- deaths (including name, sex, age and date); and
- divorce (including date of occurrence, names, date and date of marriage).

3-53 The same is the case for key variables such as citizenship (as recorded, for example on a passport), place/country of birth, and date of entry into the country and previous citizenship if applicable. Errors are mostly limited to mistakes in coding or transcription (except perhaps in cases of deliberately fraudulent reporting). In many instances such errors tend not to persist very long in a system that is frequently accessed and reviewed.

3-54 However, it should be noted that in many countries there are systematic errors in the register, that may not be easily corrected. For example, people who do not have a birth recorded, or genuinely do not know their date of birth will have a default date of birth – often recorded as 1 January for the nearest approximate year of birth. Demographic analysis by the NSO will highlight this as an error, but the administrative data holders will not necessarily be able to correct it. Where the administrative register is shown not to be perfect, the NSO will have to manage any errors, omissions or other shortcomings using statistical processes.

3-55 On the other hand, some data held in the population register, such as place of residence, are in many countries usually self-reported. While a country may impose penalties for false declarations, the data quality still depends on the diligence of individual residents if they have the opportunity and incentives to do so.

3-56 The matter of assessing the quality of population registers is discussed in more detail in Chapter 5.

3.2.2 Building and dwelling registers

3-57 Building and/or dwelling registers (referred to in some countries as real estate registers) are registers usually held by land and property valuation agencies, land survey authorities or by local or municipal authorities responsible for the development of housing policies and urban planning. With respect to a register-based population census such registers are of fundamental importance in the harmonisation of the population register, in that every person that is allocated a dwelling number can be associated with a particular housing unit, and (in some cases) a distinct household.

3-58 Such registers will often contain information on:

- the location of the building (usually as a formal address but this may also be in terms geo-coordinates),
- the type and purpose of each building (identifying in particular those that are used wholly or partially for residential purposes and whether they are occupied or vacant, or are non-residential); and
- the characteristics of these buildings that might include, for example, details of the age, layout, size and water, gas and electrical installations.

3-59 In respect of dwellings specifically, the register may include information relating to, and the characteristics of, the individual dwellings or housing units within each building, such as:

- the ownership or tenure,
- size (in terms of either useful floor space and/or number of rooms),
- floor level (in the case of apartments), and
- physical construction (such as the materials used and any technical specifications such as the degree of seismic protection).

3-60 The data for residential dwellings may be used by the NSO to create statistics relevant to the needs of a housing census. However, it may be the case that the characteristics held in the building/dwelling register can only be used as proxies for the requisite standard housing variables, such as type of dwelling, floor space, number of rooms, floor level, construction materials and period of construction. Moreover, such information will not usually relate directly to the persons living in them, and thus, if so, the NSO will need to establish a link

between the location of the dwelling and the place of usual residence in the population register. When combined with the population register in this way, the building register data can be used to provide statistics on the size of dwellings in terms of both the number of persons per dwelling and on the number of square metres per person, both important indicators to measure levels of over-crowding.

3-61 The essential statistical requirements which any building/dwelling register must fulfil are:

- The number of buildings and dwellings must be recorded in a reliable way. In particular, there must be neither double counting nor omissions in the register that would lead to over- or under-estimates of the number of buildings or dwellings.
- It is particularly important that each dwelling is assigned a unique and unambiguous identification number. Such an identification number might refer not only to a geographical reference (in terms of the administrative area of location, or the postal code, or the global positioning system coordinates, or all of these) but also, for example, to the Building Number; an Entrance Number (which is useful in the case of large apartment blocks); and the Dwelling Number within the building (which may or may not include an identification of the floor level). In some countries such identifiers remain on record and are not reallocated even after their reference building or dwelling has been demolished.
- The information should be complete, plausible and current. In other words, all required data must be available without inconsistencies and must be *updated* regularly so that, for example, the register includes newly constructed buildings, and reports when or old buildings have either been expanded or demolished. There should be administrative routines for this.

3.2.3 Enterprise (business) registers

3-62 Enterprise or business registers (sometimes referred to as ‘company registers’ or ‘establishment registers’) hold information to administer the provision of a range of services that can vary from country to country, but most

principally their aim is for taxation purposes and include registering, monitoring and storing corporate information, such as a company's legal status, its headquarters, capital and legal representatives.

3-63 In their statistical form, business registers are regularly updated, structured database of economic (or legal) units in a territorial area, maintained by the NSO, and used for statistical purposes. Statistical registers play a central role in the NSO's system of official economic statistics. Traditionally the function of the statistical business register (SBR) has been to provide a population of economic units from which frames and samples for business censuses and other economic surveys can be drawn^{23,24}.

3-64 More recently, however, the SBR fulfils two other important roles. Firstly, it is crucial in the integration and use of data from administrative and other sources. Secondly, a well-developed SBR with a comprehensive list of enterprises and other statistical units, and information about their characteristics, can be used as a source of economic statistics in its own right. Particularly for the purpose of the census, the NSO may be able to use the register to derive data on some of the core economic topics, particularly place of work and industry of occupation that may not be available from any other administrative source.

3-65 The coverage of SBRs will vary from country to country. In principle, and typically, an SBR should record all institutional units in the national economy that are engaged in economic activities. Ideally, the SBR should also include all households involved in some form of economic production. In practice, however it is impossible to identify administrative sources that cover all such small units.

3-66 The sort of characteristics that the SBR usually contains serve a number of purposes – both functional (for the purposes of managing the register and linkage with other data sources) and statistical, but only a few of which are likely to be of any practical use for within the context of the census. These would include:

- *physical address*, including an official geographic code at the most detailed level;
- *institutional sector and sub-sector*, used for distinguishing between market and non-market activities and between public and private enterprises;

- *principal and secondary economic activities*, used to assign industry in accordance with the Standard Industrial Classification at the 3-digit level; and
- the size of the workforce in terms of numbers of persons employed and employees.

3.2.4 Address registers

3-67 As noted in Chapter 2 in some countries a prerequisite for any inclusive and successful census (whether traditional or register-based) will be the availability of a comprehensive, high-quality *address register* covering all households, communal institutions and non-residential buildings. It should be noted that there are no internationally agreed formal definitions of an ‘address register’, and indeed the relationship between address and building/dwelling registers (discussed above) are country specific. Some countries may have the address as an attribute in building and dwelling registers; other countries may have the building or dwelling as an attribute of addresses in their address register; some other countries may treat each of them separately.

3-68 Whatever the national circumstances, the key factor is that a country has to be able to have a list of places where people might live, be able to assign a location, and link the place to the population register. The need to ensure that every household and housing unit is included in the census is, more than ever, a key driver in ensuring that everyone is counted in the right place, and that housing units and households are also counted in the right places. These are critical to ensure that the census statistics accurately reflect the geographic distribution of the population.

3-69 In traditional censuses, use of a comprehensive address register may enable NSOs to create lists of dwellings and households from which enumeration areas and statistical areas can be constructed and mapped so that they can provide:

- the key to accurate delivery, collection and follow-up of questionnaires;
- balanced workloads for enumerators;
- a consistently-sized output geography in which areas are not too big to mask significant local variations in character and not too small to risk statistical disclosure; and

- stratified sampling designs for post-enumeration or other inter-censal sample surveys.

3-70 However, in countries where comprehensive address lists are either incomplete or non-existent, maps created from areal or satellite imagery can often provide a workable alternative for some of these functions.

3-71 In register-based or combined censuses in which all or some of the data are derived from a range of administrative registers, an address register helps to ensure harmonisation and completeness of coverage. This becomes increasingly important as census data are produced for an ever-widening range of geographies to suit different user requirements.

3-72 The creation of such a purpose-built address list by the NSO or a government or private institution may involve the amalgamation of data from several separate and independent registers (that may have been constructed for different administrative purposes) in order to minimise under- or over-coverage. For example, lists of registered electors used for national and local voting purposes or lists of dwellings used by local authorities for assessing rateable values, may not include all postal addresses used by national or commercial mail carriers. (In England and Wales, for example, the Office for National Statistics is using a product called AddressBase as the core dataset for the address register for the 2021 Census. AddressBase is maintained by GeoPlace, a joint commercial venture between local government authorities and the Ordnance Survey (the UK's national mapping agency) drawing of data from four different sources.)

3-73 Where such a register is compiled from disparate sources (or in countries where a variety of address formats may be adopted) it should contain addresses that are in standard format that is harmonised for the purposes of the census in order to help avoid duplication issues which can occur with complex addresses or communal institutions.

3.3 Specialised or supplementary registers

3-74 In addition to the base registers described above, there will be other specialised (or supplementary/subject-matter) administrative registers available to the NSO that contain records relating to those characteristics for which census information is required and from which the requisite variables for core and non-core topics can be derived. They may also help to identify issues

of coverage by, in particular, identifying 'signs of life' activities (as discussed later in Chapter 5). Such registers may include:

3.3.1 Social security or pension registers

3-74 These are registers held by official bodies typically for the purposes of the administration of national social insurance programmes and the allocation of benefits and allowances that encompass, for example, the unemployed, families, pensioners, people undertaking parental care, and the disabled and long-term sick. The data from such registers may be used to derive census attributes for such topics as sex, age, marital status, unemployment status, income and disability/health status.

3.3.2 Tax registers

3-75 These are registers held by national and local tax authorities for the purposes of the administration and collection of income tax, purchase taxes, property and wealth taxes, building rates and other national and locally-imposed taxes. The data from such registers may be used primarily to derive census data on personal or household income and wealth that might otherwise be difficult, or too sensitive, to collect directly on a household questionnaire. Other information held on such registers may also include details of sex, age, marital status, employment status, occupation, place of work and place of usual residence.

3.3.3 Employment, unemployment and jobseeker registers

3-76 These are sources of information from which the country's official employment and unemployment figures are usually taken, although it should be recognised that the concepts of employment and unemployed adopted in national registers may not be consistent with the concepts recommended by the International Labour Organization for the purposes of producing labour force statistics. The data recorded may nevertheless enable the NSO to derive census figures relating to the key socio-economic topics of economic activity (now often referred to as labour force status), employment status, occupation, hours of work and place of work.

3.3.4 Education and student registers

3-77 These are maintained both centrally and by individual educational and academic establishments for the purpose of registering admissions and the performances of students as well as the employment of teaching staff. Such registers may also record the details of students living in halls of residence or other institutionally-registered accommodation. The data held may be used by NSOs to create census statistics on attendance, literacy and highest level of educational attainment – though it should be recognised that in some countries such available data may often only refer to the current student population. If this is the case, data on such topics with respect to persons no longer formally attending an educational establishment must therefore be obtained from other sources.

3.3.5 Health registers

3-78 These are registers (usually) maintained by locally-based health authorities for the purposes of providing health-related services, whether these are within the context of national health service or provided by insurance-based private agencies. The raw information they contain are usually treated as confidential but can be anonymised to a sufficient extent to allow them to be used by the NSO to create data on health status, domain and level of disability, and number of liveborn children. Health registers can also be used: (a) to identify population sub-groups living in health-related institutional establishments including long-term inpatients and residential student nurses; and (b) for improving the quality population registers based on ‘sign of life’ approach (see section 4.5).

3.6.6 Border control data

3-79 Section 3.2.1.6 above discusses the role of the population registers in measuring migration – and their particular shortcomings in recording international migration. Many countries, however, collect at least some data on international migration at the borders of, or points of entry into, the country, although there is considerable variation in the degree of coverage of international migration by such sources²⁵. These sources, in combination with the data held in population register (or collected in a traditional field enumeration) have the potential to create a means of investigation of the scale, patterns and impact of migration.

3-80 This is because they can provide – theoretically at least — a picture of the inflow and outflow of international migrants for a particular country over a specified period of time (one year, five years, etc.), and can include details on the type and nature of migration (such as type of visa, country of citizenship and country of origin/destination), especially if those are verified through document checks at a country's border or at a point of entry or exit. As noted above population registers alone cannot measure the quantity of movement over a particular period of time absolutely because people who emigrate may not easily be identified.

3.3.7 Other supplementary registers

3-81 NSOs may also be able to access data from a range of other administrative sources either to provide topic-related census variables or to identify particular population sub-groups that are traditionally hard to count accurately – such as the institutional population. For example (and this is by no means a complete list):

- *electoral registers* will identify households containing those residents entitled to vote (based on age and residency qualifications);
- *registers of motor vehicles* may allow the collection of data on car availability;
- *registers of foreign nationals or of residence permits* may provide information on migrants, year of entry into the country, citizenship and refugees/asylum seekers; information on refugees may, additionally, be taken from administrative records of persons in camps established to house refugees or other displaced persons;
- *lists of military service personnel* may (if access by the NSO is permitted) will indicate employment within the armed forces;
- *prison registers* can provide some basic information on members of a population group that is particularly difficult to enumerate in any type of census operation;
- *registers of building permits* can provide indicators on the extent, type and timing of future dwelling construction;

- *farm registers* will contain information relating to households with access to, or usage of, agricultural holdings that may be useful for linking the country's population and agricultural censuses; and
- *registers held by public facility service providers* may offer information on the availability of household amenities such as piped water supply, electricity and/or piped gas, and sewage and waste disposal facilities.

3.4 Options if data are not available from administrative sources

3-82 As has been made clear in the chapter, register-based censuses are generally based on a statistical population register created by the NSO from the integration of the unit record data from two or more of the types of administrative registers described above. Usually, where the census is wholly register-based, no other sources of information are used.

3-83 More often than not, however, it is neither practicable nor even possible to collect all desired census variables from registers. There will always be circumstances in which it is difficult, impossible, too costly, or even unlawful to obtain some variables from registers and where countries must rely on other data sources either to supplement such data in order to report on those census characteristics and variables that cannot be derived from the registers themselves or to improve the quality of the data obtained from administrative registers or. For example:

- Information on the inter-relationships between household members (such as cohabitation outside marriage) or even the more basic relationship to a single household reference person) that is necessary to derive household and family composition status, may not be held on any administrative register.
- Because of the subjectivity (and sensitivity) of cultural-related characteristics, such as ethnicity, religion and language, few (if any) administrative registers hold such data – indeed, in some countries it may even be unlawful for public bodies to collect or hold such information relating to identifiable individuals.

- Characteristics that relate to events that occurred abroad such highest level of foreign education obtained may not be recorded in domestic registers. Moreover, register-based information on education-related variables (such as literacy and highest level of education attained) may often relate only to those persons attending formal educational programmes.
- Variables that are based on behaviour patterns are difficult to register, such as: the *housekeeping* concept of the household definition (and where the *dwelling* concept may have to be adopted instead); and information on *mode of transportation to work* (although data from public transportation companies toll stations and mobile telephones may, in the future, be used for modelling this – see section 3.5 below).
- Administrative sources relating to the economic activity of the labour force (such as employment or business registers) will very often not hold a sufficient level of detailed information on type and place of work to enable NSOs to create occupation and industry classifications that conform to international standards;
- Key housing variables such as number of rooms or bedrooms or presence of lift or disability access may not be held on any administrative dwelling registers; and
- Situations or events occurring before the register was created are likely not be recorded, and data from registers that have only recently been created quite recently, may not be reliable since it takes time to establish sufficiently good data quality.

3-84 Where such circumstances occur, alternative sources of data might include the use of:

- existing sample surveys carried out by the NSO (such as a Labour Force Survey, Annual Population Survey, Household Income and Expenditure Survey, or International Passenger Survey);
- a purpose-built sample census or population coverage survey or other demographic survey; or
- commercial data sources, possibly including Big Data (see section 3.5below).

Or, alternatively, data relating to such elusive topics may be dropped from the census outputs altogether.

3-85 Resort to the use of such alternative sources may be particularly necessary in the case of those variables necessary for the NSO to derive the indicators by which the Government can monitor its progress towards meeting one or more of the Sustainable Developments Goals (SDGs). Of course, some of the 17 SDGs and many of the 169 accompanying targets and 230 indicators, cannot reasonably be measured by, or derived from, information collected in the census whatever methodology is adopted. For example, none of indicators needed to monitor Goal 2 (*To end hunger, achieve food security and improved nutrition, and promote sustainable agriculture*) or Goal 13 (*To take urgent action to combat climate change and its impacts*) are traditionally derived from data collected in the census.

3-86 But many of other Goals (in particular 1, 3-9, 11 and 17) can be informed, to a greater or lesser extent, by collecting information in the population and housing census³³. For example, national progress towards meeting this Goal 1 (*To end poverty in all its forms everywhere*) can be measured by four key indicators derived from information that can be collected in the census

- (a) *The proportion of the population below the international poverty line, by sex, age, employment status*
- (b) *The proportion of the population below the national poverty line, by sex and age*
- (c) *Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions*
- (d) *The proportion of the population living in household with access to basic services*

3-87 Where NSOs start to rely increasingly on administrative registers to collect its census information they will want to be assured that such data sources will contain the relevant variables so that the measurement of progress towards meeting the relevant SDGs is uninterrupted.

3-88 The range and choice of such sources is likely to differ for each country in order to comply with the essential features of the census outlined in Chapter 1. For those conducting a combined census, the key factor is the inclusion of census data derived from either a sample or full field enumeration for a selection of variables for which data is not available from registers or where the existing registers are of insufficient coverage or quality.

3-89 For example, there are no pre-existing registers of buildings and dwellings covering the whole of Germany and, as a result, the census outputs required for the 2011 housing census (in addition to some auxiliary information relating to household composition) were obtained from data collected through a postal survey. And the Dutch virtual census still relies on existing surveys to collect information on a person's occupation and industry – economic characteristics that are not available from any administrative register in the Netherlands.

3-90 Where such alternative approaches are followed, NSOs should be aware of the need to take into account the degree of error that will be introduced into the data depending on the level of any sampling adopted. This is particularly so with respect to those statistical outputs relating to the lowest level of geographic areas for which the number of collected data items is likely to be small. The potential problem of the bias that may be introduced in the linkage of data obtained from different sources also needs to be considered.

3.5 Big data

3-91 Over the last two decades more and more data have been generated on the web and produced by sensors in the ever-growing number of electronic devices. The amount of data and the frequency at which they are produced have led to the concept of “Big Data” – a term now used to describe data sets of increasing volume, velocity and variety; the three V's²⁶.

3-92 Sources described as ‘Big Data’ are often largely unstructured, meaning that they have no pre-defined data model and/or do not fit well into conventional relational databases. Nevertheless, in addition to generating new commercial opportunities in the private sector, Big Data is also of much potential interest to NSOs as an input for official statistics, either for use on its own, or in combination with more traditional data sources and administrative registers. Big Data has the potential to produce more relevant and timely statistics than traditional sources of official statistics.

3-93 Census statistics (as with most other official statistics) have traditionally been based almost exclusively on field-based data collections and/or, more recently, on the acquisition of administrative data in government programmes. But this is not the case with Big Data where most data are readily available or held by private companies. As a result, the private sector may take advantage of

the Big Data era and produce more and more statistics that attempt to compete with, if not better, official statistics on timeliness and relevance.

3-94 It is unlikely that, as a result of this, NSOs will be deprived of their 'official statistics' trademark but they could slowly lose their reputation and relevance unless they get on board. One big advantage that NSOs have is the existence of infrastructures to address the accuracy, consistency and interpretability of the statistics produced. By incorporating relevant Big Data sources into their official statistics systems, NSOs are best positioned to measure their accuracy, ensure the consistency of the whole systems of official statistics and provide interpretation while constantly working on relevance and timeliness. The role and importance of official statistics will thus be protected.

3-95 In general, big data sources can be classified as follows²⁷:

- sources arising from the administration of a programme, be it governmental or not, such as electronic medical records, hospital visits, insurance records, bank records and food banks;
- commercial or transactional sources arising from the transaction between two entities, such as credit card transactions and online transactions (including from mobile devices) for the production of economic statistics;
- sensor network sources, such as public surveillance cameras, satellite imaging, road sensors and climate sensors, for a range of geo-spatial analyses in, for example, environmental and land use studies and in the measurement of urban growth;
- tracking device sources, such tracking data from mobile telephones and the Global Positioning Systems to monitor, for example, commuter and other transportation flows and the measurement of daytime and night populations;
- behavioural data sources comprising, for example, online searches (about a product, a service or other types of information); and
- social media such as Facebook to gauge public opinion and estimate demographic and cultural profiles such as age, ethnicity and language spoken.

3-96 However, the idea of utilising Big Data is still rather novel for many statistical organizations, and there is uncertainty about what it really means for official statistics, and how best to react. Moreover, harvesting the information from Big Data and incorporating it into a statistical production process, particularly the population and housing census, presents even greater

challenges compared with using registers and administrative files held by government sources. For example:

- Big data sources can produce large amounts of timely, sometimes real-time data, usually generated at minimal cost to the provider, but which require new tools, methods and expertise to capture, manage, and process in an efficient way which may not be within the competences of the NSO and may therefore need to be done in cooperation with or outsourced to academia or the private sector.
- The use of estimation approaches may be required in the production of multivariate outputs. Different methods would be required for topics with varying degrees of missing data, as will often be the case when using multiple data sources.
- The information may not be accompanied by adequate metadata, making it more difficult to assess the quality or the completeness of the data source.
- The inherent concepts related to the files may also not be consistent with the data that is available from some registers or the information collected directly from respondents and a traditional enumeration.
- There is usually a (sometimes significant) cost involved in obtaining the data from commercial sources and the NSO would thus need to assess the cost benefit of doing so.
- And finally, in order to use such data for statistical purposes, public acceptability is needed, and a legal basis with respect to the access to, ownership of, and protection of, the data has to be developed if it is not already established.

3-97 So, in summary, the statistical community needs to adequately address issues pertaining to methodology, quality, technology, data access, data protection and privacy, management and costs, and provide a clear business case with adequate cost-benefit analyses before venturing into the relatively uncharted territory of a Big Data-based census for the future.

3-98 But any further discussion about Big Data does not fall within the remit of the rest of this handbook.

4 Transforming administrative data into census data

4.1 Introduction

4-1 The most common use of administrative registers for the purposes of a census is as a source data from which the required output variables can be derived without having to collect the relevant information directly from the public. The type, structure and content of such registers will, of course, vary from country to country depending on the administrative purposes for which the data is used by the data owners.

4-2 Chapter 3 describes the different types of administrative registers that are most commonly used by NSOs from which population base data and the requisite census characteristics may be derived. In that chapter such registers were characterised as being either:

- *base registers* - those that hold the basic information relating to the stock of the entire set of 'population' units that are being counted in the census – typically persons, dwellings and buildings, and businesses; or
- *specialised/supplementary registers* from which one or more of the characteristic variables that may be required for the census and which are not contained in the base registers (such as educational attainment, occupation, disability or income) can be derived.

4-3 This chapter discusses how the data held in the administrative registers are transformed into a statistical database from which the census outputs and analyses required by users can be produced. It covers:

- a summary of the various administrative registers likely to be used;
- creating a statistical population register;
- constructing integrated statistical registers for the purpose of census;
 - the linkage of data from different registers;
 - duplication;
 - conflict resolution;
 - updating using 'signs of life' markers;
 - editing and imputation; and
- research and testing

4.2 Creating a statistical population register

4-4 The statistical population register is a systematized and indexed collection of individual records for every resident (including nationals and foreign citizens) of the country. Where developed and functioning, it represents the backbone of the compilation of official statistics and represents a master sample frame as well for designing and running statistical surveys. Connected on a regular basis and under the provisions of the law with administrative registers, starting with the administrative population register, it is regularly updated with pertinent information. The legal framework for establishing, maintaining and exploiting the national statistical population register must ensure that it is used solely for compiling aggregate statistics and that it cannot be accessed for any other purposes within or outside of the national statistical authorities.

4-5 Establishing a national statistical population register involves extensive investment in collecting records from different existing administrative registers and/or administrative data source, harmonizing them, linking information from different registers and running complex editing procedures to ensure the consistency and quality of individual records. Once established, and under the protocols spelled out in legal provisions, the national statistical population register is updated from administrative registers on a regular basis.

4-6 Creating a Statistical Population Register (SRP) is crucial step towards achieving the use of administrative data for census purposes. It should be noted that it is important to distinguish between the registration of an individual unit (person) in the population register and the inclusion of that record in the computation of internationally comparable statistics on the stock of population and its basic characteristics, such as age and sex. While the listing of those person in the register will be in accordance with national legislation for the requisite administrative purpose, for statistical purposes only those persons meeting the requirements of being residents of the country should be included. For instance, persons who have left the country but are still holding a permanent residence permit can be still listed in the population register, but they should be excluded from the census population of usual residents. Particular care should then be applied to properly classifying registered persons who have emigrated,

especially in those countries where there is no incentive to declare the emigration to the authorities.

4-7 An SRP is usually generated from the existing population register by the NSO. In countries that have a reliable population register (for both nationals and foreigners), a statistical population register is easily generated from the existing population register. For example, in Austria, the local population registers were replaced by the Central Register of Residents (ZMR) maintained by the Ministry of Interior. On the basis of the ZMR, Statistics Austria set up a separate database, the Population Register (POPREG) which has been the main data source for producing statistics on the resident population and for producing migration statistics since then. Other registers established or in the process of being established include the register of enrolled pupils and students, the register of educational attainment, the housing register of dwelling and buildings, and the business register and their local units of employment³⁴.

4-8 On the other hand, if there is no centralized population register, a statistical population register can be constructed with integrating local population registers and linking with other existing registers. Canada, for example, used linking hundreds of administrative files obtained mainly from other federal departments. Extensive pre-processing activities took place prior to linkage to remove duplicates and standardize file variables. Given that Canadians do not possess a single unique identifier, administrative files were linked using record linkage methods; key matching variables were identified, validated and used to perform the linkage.

4-9 If the existing population register is relatively good in terms of the coverage of people living in a country but some key variables (such as place of usual residence) are not of good quality or just not available, a full field enumeration can be conducted for improving the accuracy of the stock population at local level and linking individuals to their place of residence. In this approach where the existing population register is used together with a full enumeration, data from population register may be prefilled on to the census questionnaire, and respondents may be asked to check, update and confirm their details. Other questions relating to fields not available in the register may also be canvassed during the enumeration. In the case of new individuals, households or dwellings that do not feature in the register, all fields of information that are required for the register and the census must then be canvassed afresh to create the statistical population register.

4.3 Constructing integrated statistical registers for the purpose of census

4-10 Typically a register is some sort of structured list of units, containing a number of attributes for each of those units, and having some sort of regular updating mechanism. In this way, many administrative data files can be considered to be registers, but the results of one-off data collections are not¹⁵.

4-11 A statistical register is a register that is constructed and maintained for statistical purposes, according to statistical concepts and definitions, and under the control of statisticians. Administrative registers can therefore be used as sources for statistical registers, but the reverse would normally be seen as contradicting the principle of the 'one-way flow' of data as prescribed in the Sixth Principle of Official Statistics¹⁷.

4-12 A statistical register typically plays the role of a data coordination tool, integrating data from several sources, both statistical and administrative. This may be done by linking records using common identifiers or by using matching techniques (as discussed in section 4.5 below). It may sometimes be easier to use data from a single source, but in such cases it is often difficult to check the accuracy of that source. When several sources are used and integrated within a statistical register it is possible to have a much better view of the accuracy of the data. Unfortunately, the negative side of this is that it becomes necessary to have a strategy for dealing with conflicting data from different sources. However, if variables in statistical registers are stored with source codes and dates, automated algorithms can be used to prioritise sources and resolve most data conflicts.

4-13 As well as integrating data from different sources, a statistical register may also provide the possibility of deriving new (census-related) variables not otherwise available from any one single register (such as industry of occupation, level of overcrowding or a housing quality index.)

4-14 Traditionally statistical registers have been used as sampling frames for surveys, but as has been already noted, a statistical population register is at the very heart of a register-based census. A population register may be established inside a statistical institute for statistical purposes. However, it is more common (and far simpler) for the statistical agency to transform the administrative

register into a statistical register owned and managed by the NSO, if permitted by law and the NSO has the resources to do it. This transformation process is usually a relatively simple task and may include editing, correcting, removal of duplicates, and translating the administrative ID into a statistical ID.

4-15 The NSO may also then integrate data that the administrative register does not have with authenticated evidence of data from other sources (such as, for example, country of birth for people immigrating before the administrative register was established). The statistical population register should be updated regularly. In Norway, for example, it is updated five days a week with new transactions from the administrative register.

4-16 In addition to identifying the data sources to be used for any particular statistical operation (section 4.2 notes those sources that are more commonly used in a register-based census), some of the key processes²⁸ involved in the construction of a statistical register are:

- linking the data from these sources;
- dealing with duplications;
- resolution of conflicting information relating to the same data items in different linked sources;
- updating; and
- editing and imputation.

4-17 Each of these are discussed in the following sections of this chapter. The tools for assessing the impact on data quality of each of these processes are discussed in section 5.4.

4.3.1 Data linkage

4.3.1.1 Linking data through a common identifier or alternative means

4-18 It should be remembered that it is not a necessary requirement that all the information required for a population and housing census has to be physically recorded in a single register. What is necessary, however, is the coordinated linkage of the population register with any other register containing the required information (preferably in electronic format) through the means of a common unique identification system. These other registers may, for example, be structured differently; they may report on units other than individuals; or

they may refer only to subset(s) of the population, such as the employed, students, the retired, or benefit recipients; or they may only cover local or regional areas which only when aggregated will cover the entire country.

4-19 Therefore, if data from several administrative sources are used to populate a statistical register, the national statistical organisation will need to find some way of linking those data. This will typically take the form of matching, which can be defined as the linkage of data from different sources based on common features present in those sources¹⁵.

4-20 This process of record linkage can be done by matching individuals based on name, sex, date of birth, and address. But in practice this is complicated and time-consuming and far from perfect because many people may share some of the same characteristics and/or because some of the data may be missing, incomplete or erroneous for some of the variables. The assignment of a personal or unique identification number (PIN or UIN) to everybody, and the use of these in administrative and statistical registers, makes the matching significantly easier, faster and less costly.

4-21 If these common features include some sort of reference or identification number (referred to as a common ID), the process can be referred to as 'exact matching', and is relatively easy. In exact matching there are two possible outcomes, either two records from different sources match exactly on the basis of the common IDs used, or they do not. In other words, a record with the ID number 123456 will match to a record relating to the same unit in a different source with the same ID, whereas it will not match to a unit with the ID number 123457.

4-22 Exact matching relies on the quality of the matching variables used in each source. If there are errors in the common IDs in at least one source, there is a high risk of either matching the wrong units, or failing to match units that should be matched. For this reason, even when common IDs exist in all the files to be matched, it may not be sufficient to rely on exact matching alone.

4-23 Sometimes identifiers can include check digits, that is one or more characters that are generated according to a standard algorithm based on the other digits in the identifier. If check digits are present, they should help to guarantee a certain level of quality by eliminating most typing or reading errors.

4-24 A method to increase personal data protection is to encrypt the ID numbers in different registers to prevent information from being read by unauthorised parties, and where matching can only be done by decryption.

4.3.1.2 Identification numbers

4-25 As noted above, a pre-requisite for implementing a register-based census is thus the ability to link accurately unit record data from different administrative sources that refer to the same enumeration unit, usually an individual person, household or dwelling. In most countries this can be achieved by means of a unique identifier that is common across all the relevant registers. There is likely to be a significant difference in the accuracy of the linkage between administrative sources – and hence the quality of the census data – where a unique identifier is employed compared to those countries where no such key is available.

4-26 In some countries ID numbers have been introduced for statistical purposes but more often they have been created for administrative purposes to be used in population registers, civil registers, national identification systems or other administrative registers for the purposes of, for example, work, taxation, receipt of benefits, health care, educational services or other government-related functions including the issue of passports or other ID cards or identity documents. In such cases the design and use of such numbers are likely to have been unique to each specific purpose and not adopted for wider and more general usage. In Norway, for example, the impetus came from employers in order to facilitate the registration of employees for tax purposes, while in GCC countries, employment is linked to the ability to reside in the country, and a national identification system is of fundamental importance to managing this link. In other countries - South Korea for example - the administrative ID number is transformed into a statistical ID number. Moreover, some countries use different ID numbers in different administrative (or statistical) registers. If links between the different number series are known, records from different registers may be matched. (See also Box A for more examples.)

4-27 The ways in which such a number is implemented vary among countries, but in many cases, citizens are issued an identification number upon reaching legal age, or when they are born. Persons may be registered as non-citizens and issued such numbers when they enter the country, or when granted a temporary or permanent residence and/or work permit, or at birth if born to non-citizen parents.

Box A National Identity numbers

National Identity (ID) numbers – more often referred to as Personal Identification Numbers (PINs) - are commonly used by the NSO in many countries as a means linking the information held in different registers that relate to the same unique person for statistical purposes. They will, in most cases be the same ID that is generally used by national governments for purposes of tracking their citizens, permanent residents, and temporary residents for the purposes of work, taxation, receipt of benefits, health care, and other governmentally-related functions including the issue of passports or other ID cards or identity documents .

The ways in which such a system is implemented vary among countries, but in most cases, citizens are issued an identification number upon reaching legal age, or when they are born. Persons may be registered as non-citizens and issued such numbers when they enter the country, or when granted a temporary or permanent residence and/or work permit.

Many countries issue such numbers for a singular purpose, but over time, they become a *de facto* national identification number. For example, the United States developed its social security number system as a means of identifying and paying social security benefits. However, due to function creep, the number has become used for other purposes to the point where it is almost essential to have one for the purposes of for example, opening a bank account, driving a car, or to obtain a credit card.

4-28 There are two different principles for the design of a PIN. The first is where the PIN comprises digits that relate to date of birth and sex, often place of birth or registration, and sometimes on citizenship or name. This is the case in the Nordic and many other countries. The advantages of this are that it is relatively easy to remember the number and to spot errors. The disadvantages are that it may reveal personal information that is regarded as sensitive, and that it cannot be accurately determined if the date of birth is not known. Inclusion of place (or country) of birth in the PIN may under special circumstances facilitate discrimination of persons born in certain areas (or countries). Moreover, the whole number has to be changed if it is later found that the date of birth is wrong. It also happens, although rarely, that a person changes sex and consequently needs a new PIN.

4-29 The second case is where the PIN is a unique *information-free* random or serial number, where the digits have no specific meaning. The advantage of this is that privacy is better protected. In addition, information-free numbers require fewer digits than those that include date of birth, etc. All required information about a person can be stored in a database, typically the population register,

where it is available for legitimate use. Information-free numbers are less prone for misuse, since information about a person cannot be detected from the number alone, but such numbers may be more difficult to remember.

4.3.1.3 Linking datasets without a common ID

4-30 There is likely to be a significant difference in the quality of the linkage between different administrative data sources with and without common IDs. If there is a unique identifier in all or most of the records, the linkage becomes relatively easy and the level of successful linkage is normally high (although the quality of the variables in the statistical register should still be assessed).

4-31 However, the NSO will need to develop alternative methods to link data where in those censuses where either:

- (a) the unit record data from registers are being combined with the information that is collected directly from individuals and households in a field enumeration (as in the case of a combined census) and where ID number is not reported on the questionnaire: or
- (b) there is a lack of any unique identifiers across administrative data sources (as in the United Kingdom and New Zealand for example).

4-32 Such methods may involve either:

- *Deterministic method* in which a number of match-keys (such as name, sex, date of birth and postcode) are used to link records across the administrative sources. The same set of match-keys is produced for each dataset. If the match-keys are the same on each source, a link is made - however, the link may not be unique.
- *Probabilistic method* that identifies links between records in two datasets by comparing and quantifying the relative similarity of records and assigning a 'similarity score' – a process that is sometimes referred to as 'fuzzy matching'. This has the benefit that, compared to the deterministic method, it does not require record values to be identical in both records). However, the disadvantage here is that some links may be wrong – but this may have little effect on the overall statistical results.

4.3.1.4 Linking persons to dwellings

4-33 The basic counting units of a population and housing census include not only persons, but also households, families, and dwellings (whether occupied or vacant). All of these units require identification, but there is no need to use all the different identification variables. The minimal necessary identifiers are for persons (person ID) and dwellings (dwelling ID).

4-34 These IDs must be linked with each other, so that a dwelling ID is assigned to each person. This ensures that each occupied dwelling is associated with a list of person IDs of all of its occupants. The dwelling ID makes use of the address code, which may contain also spatial coordinates.

4-35 For those countries conducting a traditional census, information about households is usually collected on the basis of the *housekeeping concept*. This definition can be achieved through asking a specific question, but is more challenging for countries conducting a register-based census. Many such countries instead use the *household dwelling concept*, which considers all persons living in the same housing unit to be members of the same household. While adopting this definition has minimal impact on the total number of private households, it can have a larger impact for certain household types, such as one-person households. This bias in the number of private households and in the estimated structure of the household types depends on the traditions of the country and on living conditions. These challenges for register-based censuses also extend to the construction of families within households using relationship information.

4-36 In some countries (such as Slovenia) a distinct household register exists. The existence of such a register eases the organisation of a register-based census, especially when household IDs are included in the register, when there is then accurate information available about which person ID belongs to which household ID. A household register might therefore improve the quality of a register-based or combined census significantly. However, the situation of Slovenia is not commonly adopted.

4-37 Sometimes it is useful to use identification numbers for other units such as enterprises and organisations. If these are linked with person IDs and dwelling IDs they form a helpful tool for deriving other census characteristics, such as commuting patterns between place of residence and place of work.

4.3.2 Dealing with duplication in the statistical population register

4-38 Duplication of persons in the statistical population register can result, for example, from loading information for a given person from multiple data sources that do not carry the same unit identifier. If good record linkage methods are not in place, information from a second data source may not be associated with the person present in the statistical population register and thus results in creation of a new person that duplicates the existing person.

4-39 Duplications can be removed through a record linking/matching process if all units in the statistical population register are matched with one other. This search for, and the elimination of, duplicated persons ensures that newly created units are true births or immigrants, and deleted persons are true deaths or emigrants. This is essential in constructing longitudinal data. Skipping this step may lead to the duplication of persons and create coverage error .

4.3.3 Conflict resolution

4-40 As noted at section 4.3.1 above, when combining administrative data to create statistical registers there may be inconsistencies in the values of key variables across different sources. For example, once a decision has been made on which administrative records to include for the usually resident population, if their address on two or more sources are different (a common occurrence where there has been a delay in individuals communicating a change of address, or where there have been administrative processing delays, or in the incidence of second/multiple homes), then the NSO may need to decide at which single address the records should refer. Conflicting (or multiple) address information and any related decision may cause under-coverage in some areas and over-coverage in others. When aggregated at the national level this may not be an issue because whichever address is reported, individuals and households may only be counted once; but at the sub-national level, and particularly so at the small-area level, this may matter, if the two addresses are located in two different areas, as it will cause over-coverage in one area and under-coverage in the other.

4-41 With this particular example of conflict in mind, a common approach is to choose which source is considered to be the most likely to be up to date based on the characteristics of the individual or the administrative variables (this

approach could also include using additional data sources where the same individual appears).

4-42 Similar approaches in using quality information/indicators on individual sources to measure the quality of attributes in statistical registers, when the same attribute is available in different sources, have been used in other UNECE countries. For example, the Spanish population register lacks information on the legal marital status (LMS) of individuals. To estimate LMS, therefore, several registers are used to obtain complete information, including data from the Tax Agency, the Civil Register, the Social Security database and the Central Register of Foreign Nationals. Since an individual may appear in multiple data sources with conflicting information, decision rules are applied to determine the most plausible value. The decision rules are applied for each person after which a value for LMS may be given. If cases remain unassigned, a value is imputed depending on age, information in past censuses, and number of household members. The results generated by this method are promising and further research is on-going³⁴.

4-43 Possible solutions to resolving other areas of potential conflict associated, for example, with differences in definitions, classifications, reference periods and other inconsistencies between sources, are described in section 2.4.

4.3.4 Updating and the 'Signs of Life' methodology

4-44 The basic minimum input for keeping population registers up-to-date is the information obtained from the civil registration of births, deaths and marriages, and of any changes of address resulting from either internal or international migration. However, an increasingly commonly used tool to help minimise the almost inevitable shortcomings in coverage in the statistical register is the so-called 'signs of life' (SOL) method in which a number of 'activity rules' are adopted to ensure that only persons who are alive and meet a set of pre-defined residency criteria are included in the census. The data from those administrative data sources to which the NSO has access can be cross-checked to improve the coverage of statistics derived from a population register by using the information from the different registers that relate to the activity (signs of life) of the same resident at a particular period in time. Methods of assessing the quality of the registers are described in more detail in Chapter 5.

4-45 Based on such information it is possible to update the statistical register so that revised estimates of the size, migration, household structure and spatial distribution of the population in the country can be made. As an example, when using household data from registers (where a household equates to those persons living in the same dwelling), it may be that the distribution of different household types seems to be biased when comparing it with earlier census data or data from household surveys. In this case, it may not be necessary to recalculate all families, households and their dwellings, but only correct the seemingly biased results. To improve this bias, the signs of life markers from different data sources maybe used.

4-46 A list of such markers can never be absolute but the more markers that can be used (such as registering with a doctor, enrolling in course of formal education, or receiving any social support such as unemployment benefit) the more accurate will be the judgment (with lowest probability of error) that a revision to the statistical register should be made. In making such statistical decisions, it is useful to use the information about markers relating not only to the most recent situation but also to previous activities.

4-47 To use these markers two key pre-conditions need to be satisfied, at least at the national level:

- (a) a set of administrative registers should be held in which,
 - all persons in all registers are identified by their unique ID codes; and
 - all living quarters (dwellings, family houses, etc.) in all registers are identified by their unique address IDs;
- (b) all registers should cover the whole population and be regularly updated (at least annually)

4-48 In Spain, for example, such signs of life are determined by using data from a number of administrative sources (including tax registers, social security files and the national unemployment database) together with movements detected within the population register, the *Padrón*. These administrative sources are linked at the individual and household level, and persons who exceed a particular SOL threshold level, indicating an 'active' presence in the country, are included in the population count, while all the others (the 'inactive') are excluded. A short case study relating to the use of signs of like markers (and which has been previously illustrated by UNECE²⁸) is set out in Box B below.

4-49 Though the use of existing SOL markers in a data source may allow the optimal use of data already held, NSOs should investigate carefully the actual content and coverage of the markers from a national perspective to ensure consistency and harmonisation, particularly in respect of frequency, timeliness and the data collection methodology.

4-50 The use of markers from administrative data sources gives a different view on known datasets and allows a comparison of multiple sources. However, identifying events that are systematically missed data when no comparison is available remains challenging.

Box B Signs of life in administrative data sources: Spanish case study

To identify which individuals are usually resident, the ‘signs of life’ (SOL) method is applied. All individuals are analyzed within the available administrative data sources and the movements are detected in the *Padrón* for the months following the reference date. The four key administrative sources used to assess SOL are:

1. Tax Agency and local tax files.
2. Social Security Insurance Database that includes individuals with insurance and beneficiaries (employees and pensioners).
3. Labour market-related sources including:
 - a. Unemployment National Service Database that provides a job seekers file to include individuals unemployed.
 - b. Social Security Affiliation Registers that provide affiliation information of the employed population.
 - c. Public Aids Database that provides information about benefits recipients.
4. Central Registry for Foreign Nationals Database that provides supplementary information about foreign nationals living in Spain such as date of application for residence permit, licence or rejection of residence permit, expiration dates residence checks etc.

Through using the SOL method, individuals who reach the threshold of presence signals within administrative data will be identified as ‘active’ and will be included in the population counts. In addition, individuals not meeting the threshold will appear ‘inactive’ and will not be included. These SOL from administrative data can also be compiled at individual and household level, therefore information is available about how many household members are ‘active’.

Furthermore, for both Spanish and foreign nationals, the movements in the *Padrón* are taken into account in the following months after the reference date. There are certain movements that require the direct intervention of the person, or a residence check made by a municipality, which generates a high probability of the person to be residing in Spain at the reference date. Also, other movements are good indicators of a person not residing in Spain at the reference date. These movements can be used to identify individuals that are ‘usually resident’.

For children, it is considered to be a sign of life if an adult in the same household shows a sign of life. Children who do not meet this requirement are currently excluded from the population, though the possibility of using information about enrolled students in official studies is currently being analysed.

4.3.5 Editing and imputation

4-51 The editing of administrative-based data is usually the last step in the process required to produce a statistical register. The assessment of quality of the source and input data will inform whether or not the administrative data requires editing (to correct values that are either clearly erroneous or implausible) and/or imputation (to insert plausible values where data items are missing). Editing and imputation may be required both on any single source and the integrated database.

4-52 One of the most important and resource-consuming processes in the creation of statistical registers is the verification of the basic data taken from the administrative data sources and the correction of any errors (through editing) and the replacement of any missing values (through imputation). However, it should be noted that unlike the information recorded on questionnaires in a traditional data collection operation, in a register-based census any one particular data source should have already been checked and edited (to some degree) by the data holder in order to fulfil the administrative purposes for which the register itself is used.

4-53 This means that the focus of any such checks will be on the variables that are most important for the work of the administration. As noted by UNECE⁵ the data that are to be used subsequently for statistical purposes may not be controlled as carefully. It is vital therefore that the NSO has good knowledge of the metadata of these editing procedures (as noted at section 2.2.6) in order to fully understand any systematic limitations. Thus, it is the function of the NSO to edit the data to the extent necessary to ensure that they are of sufficient quality (in terms of accuracy) for statistical use.

4-54 The metadata that accompanies the supply of the data and/or discussion with the register holders should provide an indication of degree of such editing. Thereafter, the NSO will need to check and, if necessary, edit the data further in order that it is fit for the purposes of the census – or any other statistical operation undertaken by the NSO. (See the UN's *Handbook on Population and Housing Census Editing, Revision 2*⁹ for a general discussion on data editing within the context of a traditional census.)

4-55 Where data from several sources are edited simultaneously, consistency checks can discover further errors or inconsistencies in any of the sources that may not be evident from a single register alone, since the coverage of different

sources will often vary. Such ‘consistency’ checks are usually not applicable in traditional censuses where only one source of data is used, but are a key component of the editing process in a register-based census as they can identify errors not only in attribute data items but also in the statistical units themselves.

4-56 The statistical register should ideally be clean and consistent, so that there are no conflicts between individual data items. Improbable and/or missing values information must be investigated. In this process it is not unusual to discover systematic errors in the administrative registers, though these should, preferably have been identified prior to data processing in the extracts that the data supplier makes before the NSO receives the full dataset (see section 5.3.1 in the next chapter).

4-57 Within the context of a traditional census (or a combined census) where all (or some) of data is collected by means of a field enumeration or survey, errors are normally identified as being *coverage errors* or *content errors*, where:

- *Coverage errors* arise from omissions or duplications of persons or housing units in the census enumeration. The causes of coverage error may arise from:
 - incomplete or inaccurate maps or lists of enumeration areas or living quarters, or
 - the failure of enumerators to canvass or interview all the units in their assignment areas or where such units are enumerated more than once;
 - the omission of persons who are unwilling to be enumerated or are hard to find;
 - the erroneous treatment of persons whose residency status may be difficult to determine, such as visitors and non-resident foreigners; or
 - loss or destruction of census data prior to data processing;

and where:

- *Content errors* arise from the incorrect reporting or recording of the characteristics of persons, households and housing units. Such errors may be caused by:
 - poorly designed questions or poor sequencing of the questions,
 - poor communication between respondent and enumerator, leading to either enumerator or respondent error; or
 - mistakes in coding and data entry.

4-58 Where census data are taken from administrative registers many of these causes of error are, to a greater extent, eliminated but not - it should be noted – entirely. Errors may continue to consist of a failure in the information to meet the statistical requirements for the purposes of the census. It is crucial that any systematic types of error occurring in the administrative registers from which data are extracted for statistical purposes should be corrected. This can only be achieved by close contact with the authorities responsible for the register. Consequent improvements in the way that data are collected and correctly recorded in the administrative register can only be to the benefit of the register holder, the NSO and the data subject⁵.

4-59 Where isolated errors are detected, they may in certain cases be corrected by direct reference to the source registers. However, in practice, as most administrative data sources cover the total population, the possibilities for the data holders to contact individual units to correct the original information are very limited (and in any case would be very time consuming). This is particularly the case when the units are individual persons. The main method for control and corrections is therefore computerised, logical editing.

4-60 To adjust the statistical register to minimise the effect of missing data, the missing values can be replaced (or imputed) with synthetic but plausible values that are formed either by:

- statistical estimation (such as hot-deck or logical regression) using one or more probability distributions;
- using a deterministic model exploiting the relationship between the unknown value and other known variables; or
- some form of statistical matching.

4-61 Whatever imputation process is employed, the aim should be to fulfil a number of quality-related criteria, notably:

- predictive accuracy by which the imputed values should be as ‘close’ as possible to the true values;
- ranking accuracy by which the order of imputed values (for attributes which are at least ordinal) should be preserved;
- distributional accuracy by which the distribution of the true data values should be preserved; and
- the plausibility of the imputed values.

4.4 Research and testing

4-62 It is inadvisable for new administrative data sources to be integrated into census production without prior feasibility research by NSOs. The quality of a data source may be established by acquiring ‘test data’ and assessing its quality at the various stages suggested in this handbook. This will aid the design of a census methodology that makes the most of the available administrative data and considers the impact of its use on the quality of the census overall.

4-63 Firstly, feasibility research involves developing a detailed understanding of the administrative authority’s data collection processes, the population covered, and variables included within the source as well as how accessible this data is (see section 5.3.1 for a more detailed discussion). Secondly, supply, acquisition and ingestion of test data should be rehearsed, and test data examined in detail to identify quality issues and define cleaning and harmonization, along with validation checks (discussed in section 5.3.2).

4-64 When data from multiple registers are combined, they can be used for verifying data quality on the one hand, and, on the other, for selecting the most reliable variables and values, in accordance with the developed methodological rules. Finally, estimates produced using test data can be compared with previous census estimates or any other such source that the NSO considers to be a ‘gold standard’, contributing to an assessment of the overall quality of the output.

4-65 Generally, census characteristics cannot be acquired directly from administrative data sources, because they have been designed for other, non-statistical purposes and thus most of the definitions and classifications used by administrative authorities are different from standard statistical definitions. As such, data from multiple registers may be used in order to construct or derive certain census characteristics, while other characteristics may be covered by duplicate information in several registers. This makes feasibility research a key stage for developing methods for the derivation of census characteristics.

4-66 NSOs should address the following main challenges when deriving census characteristics:

- ascertain the international census standard (definition, classification, etc.) applicable to the target census characteristic;
- compare and contrast census definitions and classifications with the definitions and classifications used in the administrative source;

- test the accuracy of the administrative data recorded against alternative sources and work collaboratively with data suppliers to eliminate/mitigate any shortcomings;
- determine which and how many sources are required to derive and quality assure each target census characteristic;
- establish optimal rules for deriving each census characteristic and develop the necessary data processing software, optimised for the quality of outputs sought; and
- where characteristics are not covered by any administrative sources, take steps to ensure creation of the necessary register or register part (e.g. suggest amendments in register procedures, the legal environment, etc.).

4-67 Reaching an adequate harmonization of register and census concepts can be a complex and time-consuming activity which should not be underestimated. As such, it is recommended that the inclusion of administrative data sources into census production should be preceded by adequately resourced feasibility research which provides a ‘proof of concept’ for the planned integration of administrative data into the census production.

4-68 In addition, making the four stages of quality assessment (to be discussed in Chapter 5) an integral part of feasibility research will enable NSOs to directly apply the learning from such research to the census production context and to better inform users on the quality of data sources (as noted at section 5.4 (Reporting quality)).

5 Quality assessment

5.1 Introduction

5-1 Census outputs are produced for a wide range of users who will have well-established, but often, different levels of expertise, understanding and expectations. Therefore, in transitioning to a register-based (or combined) census methodology, NSOs should establish a high level of user acceptability of the outputs that such a census can offer. The quality of the census data should at least be as good as that achieved in previous censuses, while the use of administrative data can provide the opportunity to increase the range and periodicity of statistical outputs. In particular, a big advantage of register-based data is that they offer information of the entire target population on a regular basis (annually or in some cases even more frequently). This means that a register-based system is very suitable for producing statistics on small population groups and for small geographic areas more frequently than is possible with a traditional census, subject, of course, to the usual requirements to protect confidentiality.

5-2 However, as a result of the transition, there may be some changes to definitions of variables, population bases and output classifications. The impact of these changes on the quality of statistical outputs should be investigated and assessed by means of thorough testing and evaluation programme (as noted in Chapter 4), and the outcomes explained to users.

5-3 It is often the case, however, that assessing the quality of the statistics produced using administrative sources presents a particular challenge due to the complexity and multi-dimensionality of the data used: indeed, many factors affecting quality are not quantitatively measurable. Moreover, being 'fit for purpose', means, by definition, that what high quality statistics should look like will necessarily vary from one user to another; for example, while some users will regard accuracy as being of prime importance, others may prioritise timeliness and accessibility.

5-4 Regardless of the data collection methodology, assessing the quality of the output of census data has always been an important and necessary task. There are several different ways and methods to assess the quality of statistics and particularly the quality of census outputs. It is always useful if more than one method is adopted.

5-5 This chapter discusses how NSOs can assess the quality of data obtained from registers with the aim of producing good quality statistical outputs as measured against a number of standard quality-related dimensions. It begins by reviewing the key dimensions of quality, before describing how these dimensions can be assessed in the four stages of quality assessment in the census process (*source quality, input quality, process quality and output quality*). The key tools and indicators that can be utilised to quantify the assessment of quality at each of these stages are then discussed.

5-6 More detailed information is available from the recently published UNECE guidelines for assessing the quality of administrative sources for use in censuses²⁸ (from which much of the material in this chapter has been taken) and from the second revision of the quality assurance framework of the European Statistical System³⁶.

5.2 Quality dimensions

5-7 There are several internationally accepted dimensions that can be applied to aid NSOs in their assessment of the quality of census data and processes to enable both producers and users to understand the quality of a statistical output. These key dimensions are (in no definitive order of priority):

- relevance,
- accuracy and reliability,
- timeliness and punctuality,
- coherence and comparability, and
- accessibility and interpretability.

5.2.1 Relevance

5-8 ‘*Relevance*’ is defined as *the degree to which statistical outputs meet current and potential user needs*. It should be noted therefore that a census programme should maintain a close relationship with the users of census information, and it is one of the crucial steps in the preparations for any census. This dimension may require NSOs to adjust the direction of their output programmes over time, as user requirements change. However, assessing relevance is subjective because it often depends on varying user needs. The

challenge for any census programme is to balance as evenly as possible the, often conflicting, requirements of different census use while keeping within tight resource constraints.

5-9 To be deemed relevant, census outputs should, for example, not only contain the information sought about the target statistical units, for the target reference period, but also be congruent with the reasons for transitioning to the new methodology (such as reducing costs and response burden, etc.).

5-10 Users' requirements for data should be determined through consultations early in the planning phase of the census operation as they are key in shaping the design and content of the census. Such requirements will, in turn, determine the particular registers that the NSO will want to use.

5-11 In transitioning to a register-based census it is particularly important that users should be aware of any changes both to the definitions of variables and to the output classifications. It will be important that users should understand that administrative data have not (in general) been collected for statistical purposes and that the relevance (and indeed several of the other quality dimensions) of the data will be highly dependent on the importance of the data to the administrative authority's function.

5-12 As a consequence, a register-based census may:

- not always have the same range of variables where the data is not available in any administrative source;
- have to adopt different concepts and definitions, particularly in the way that different populations bases (such as usual resident population and the labour force) are defined and counted;
- have to manage sources with different reference periods;
- have limited coverage of the target population for particular sub-groups of the population; and
- also be subject to changes over time and inconsistencies in the way the data are collected across units of the population.

Therefore, there may be disruptions in any time series that span the period of transition, the effect of which will need to be communicated to users.

5-13 The impact of these factors on the quality of statistical outputs should be investigated and assessed by means of thorough testing and evaluation programme, and the outcomes explained to users.

5-14 Cognitive research within focus groups, and other tools can be used to understand the range of administrative registers that will be necessary to satisfy users' needs. Though there will likely be a set of core topics on which NSOs will want to collect information - either to fulfil international obligations (such as the EU Census Regulations) or to conform to the UN's global recommendations - the use of some administrative registers (or other data sources) may need to be assessed to meet purely national and local requirements.

5.2.2 Accuracy and reliability

5-15 The '*accuracy*' of statistical information is defined as *the degree to which the information correctly describes the phenomena it was designed to measure*. It is determined by the closeness between an estimated result and the unknown true value. It is usually characterized in terms of error in statistical estimates and is traditionally broken down into bias and variance. However, that latter concept is relevant only in situations where (in a combined census) some of the data are collected for a sample of persons or households, or where only a sample of records is processed.

5-16 In several countries the overall accuracy of census data has traditionally been based on dual system estimation (DSE) which involves conducting an independent large sample post-enumeration survey immediately after the census data collection operation. The process then relies on a capture-recapture methodology to estimate *coverage* errors (due to either under- or over-coverage) and *content* errors (arising from any misreporting or mis-recording of data items). (More details on the methodology and processes of post-enumeration surveys can be found in the UN *Technical Report on Post Enumeration Survey Operational Guidelines*²⁵.)

5-17 However, such a method may not always be appropriate or applicable in a wholly register-based statistical system in which such an independent field operation is not practicable, and therefore new or revised methods are needed for register-based censuses. An example of this is the methods that some countries have developed to reduce over-coverage, through the 'signs of life' methodology, discussed at section 5.3.3.2 below.

5-18 ‘Reliability’ is the degree of closeness of *initial* estimates to *subsequent* estimated values measured over time and different geographies. This dimension is also closely associated with ‘comparability’ (as discussed at section 5.2.4 below).

5.2.3 Timeliness and punctuality

5-19 ‘Timeliness’ generally refers to *the delay between the date to which the date refers* (usually referred to as census day) *and the date on which the information becomes available*. In the context of a register-based census, timeliness of the data can also be taken to refer to the length of time between the date of the event recorded in the register and the census reference date, or the length of time between the most recent updating and when the data are delivered to the NSO. By analogy, ‘punctuality’ refers to *degree to which the actual delivery of the administrative data to the NSO differs from the date that has been pre-agreed between the NSO and the data administrator*.

5-20 Often in any census output timetable there are several release dates to be considered. Typically, there is a trade-off between timeliness and accuracy in that the more attention that is given during data processing to producing accurate results the longer that processing usually takes. In some cases (and depending on the users’ primer needs) it may be preferable for the NSO to wait to get higher-quality input data, or a second data supply for the same reference date. Timeliness can also have impact on relevance of the data to users.

5-21 A number of other key factors should be considered when assessing timeliness and punctuality. These are discussed at paragraph 5-42 below.

5.2.4 Coherence and comparability

5-22 ‘Coherence’ refers to the *degree to which data that are derived from different sources or methods, but which refer to the same phenomenon, are similar*. More specifically, within the context of the published results, it also refers to *the degree of consistency in the value of a count or estimate where this appears in different outputs* (see section 5.3.4 below).

5-23 ‘Comparability’ can be seen as a special case of coherence that refers to *the degree to which data can be compared over time and domain*; it is closely associated with the dimension of ‘reliability’ as noted at section 5.2.2 above).

5-24 By its nature, a census that uses various administrative data sources will face the challenge of maintaining coherence and comparability. The key factors to be considered are related to:

- *Differences in concepts*: The variables in the administrative source may use somewhat different concepts than required by the NSO (such as the definition of ‘usual residence’ or a ‘household’).
- *Differences in reporting methods*: For example, education or employment status recorded in the census using register-based data may give different estimates than those in reported directly by the householder in field enumerations or sample surveys.
- *Change in methods over time*: The move to using (more) administrative sources in itself creates a change in data collection methods that is likely to affect comparability.
- *Changes in the administrative sources*: Changes in the administrative sources over time should be carefully monitored in order to avoid, or minimize, negative impacts on coherence and comparability (such as changes in the legal or administrative requirements concerning the collection of the data by the administration, or changes in the processing by the administration).

5.2.5 Accessibility and interpretability

5-25 ‘Accessibility’ is defined generally *as the ease with which users are able to access census data*, including the ease with which the existence of the data and metadata can be ascertained by the user as well as the suitability of the format and/or medium through which this information can be obtained. In the production process this definition can be applied to the input datasets from which the output statistics are created. Additionally – and specifically within the context of a register-based census – it can also refer to *the ease with which the NSO can obtain the input data in their entirety from the data holder*.

5-26 ‘Interpretability’ (sometimes referred to as ‘clarity’) relates to *the availability of any supplementary information or metadata that may be necessary of help the user to interpret, understand and utilise the published data*.

5-27 Some key factors in the context of using administrative data for the census are:

- *Availability of metadata concerning the administrative sources:* If the metadata of an administrative source is not available, or is insufficiently detailed, the usefulness of that source for its intended use (by both the NSO and the end user) will be reduced.
- *Proper documentation of the production process:* Not maintaining and updating the metadata at each stage of the production process, especially when several sources are combined, will reduce the clarity of outputs.
- *Complexity:* Special attention should be paid to explaining the complexity of the sources and of the production process to the user. There is often a trade-off between the number and complexity of sources to be used and the difficulty of understanding by users of how the final output was produced and how good it is. Indeed, the complexity of the administrative data and the availability and completeness of the associated metadata will impact on the ability of the NSO itself to understand, access and use an administrative source. For example, administrative data can be held in large, complex data structures, posing significant technical challenges for the NSO to overcome.

5.3 Stages of quality assessment

5-28 The assessment of the quality of administrative registers and the census data derived from them can be carried out in four stages, broadly corresponding to the stages of the statistical processes of the census: data collection, data processing and output production.

5-29 While the statistical design of any census is never entirely linear (meaning that the stages run consecutively) considering the stages of quality assessment in this way should enable NSOs to identify the key quality considerations at each stage of the census production cycle. These four assessment stages refer to:

- *source quality:* a metadata-based quality assessment of new or existing administrative data sources to be used in the census;
- *input quality:* an assessment of the raw administrative data as it is supplied to NSOs by the administrative authorities;
- *process quality:* the assessment of changes in the quality of the data which results from integration and processing of the administrative data within the NSO;

- *output quality*: the overall quality assessment of the statistical results as disseminated to users.

5.30 It should be noted that the information and insight gained through the assessment of the quality of administrative sources and the input data is useful not only to determine whether a particular source should be used in the census, but also to determine the necessary processing of the administrative data for use in a census. As has been previously noted, administrative data cannot be usually used directly in a census, due to conceptual and definitional differences and due to limitations of coverage, completeness and accuracy. It is therefore necessary to transform the data taken from administrative sources using the information gained at the Source and Input data stages.

5-31 The following sections of the handbook discuss, for each of these stages: the relevant key quality dimensions; the means of assessing quality and the criteria against which such assessment may be carried out; and a summary of the key tools and indicators that can be applied in such assessments. A fuller description of these tools and indicators is set out in the recently published *UNECE Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses*²⁸.

5.3.1 Source quality

5-32 In the transition to a register-based (or combined) census methodology from a traditional approach, it is necessary to assess the quality of the administrative registers at the earliest possible stage. Indeed, as noted in Chapter 2, it is a pre-requisite that the NSO should undertake research into the feasibility of using any register prior to the completion of the census design. This stage does not necessarily require NSOs to be in possession of the actual data items but is critical for the stages that follow.

5-33 But even where data from a particular register has been used before, further re-assessment should be made prior to each census as the data might have since been updated and suffered changes in terms of concepts, classifications, coverage and collection methods. Remember, apart from their use in the census, the NSO may also wish to use such administrative data sources to produce other types of outputs as part of a general register-based statistical system.

5-34 The methods of assessing the quality of data sources maybe grouped in four main categories:

- *Management and planning:* This covers those activities that relate to relationship between the NSO and the data suppliers. It requires the establishment of governance and rules, maintaining good communication, and the management of agreements, policies, and delivery plans. Any data exchange agreements will need consider not only any relevant legislation and confidentiality constraints but also level of public approval of the use of a particular data source.
- *Organization and responsibilities:* This covers the aspects associated with the establishment of responsibilities of working with data sources. In particular, it is important that the supply of data from different providers is coordinated to ensure a timely delivery.
- *Processes:* This deals primarily with the systems that provide for the transfer of (usually large volumes of) data, for which exchange systems have to be established and tested. Data transfer has to be managed in a secure environment that provides the necessary levels of privacy and confidentiality.
- *Tools:* This covers those aspects associated with adoption and development of tools to support data source quality. Standards for data formats, data models, data transfer protocols, established documentation of the administrative sources (metadata) describing the objects (units, events) and variables with their classification. A privacy impact assessment is, for example, a tool that relevant for assessing public acceptability, and a legal framework is necessary for ensuring the authority of the NSO to acquire and use the data.

5.35 All but one of the quality dimensions described above are pertinent in the assessment at this stage (relevance, timeliness, coherence and accessibility) plus an additional key element – the institutional environment – which is discussed below. The quantitative dimension of accuracy refers principally to the data items themselves rather than to the quality of the register as a source. This dimension comes into its own later in the discussion on input quality at section 5.3.2.

5.3.1.1 Relevance

5-36 An understanding of the differences between the administrative population and the required census population, and between the

measures/variables in the administrative source and the required census characteristics is important in assessing relevance. The errors arising from these differences can be referred to as 'representation' and 'measurement' errors respectively.

Representation errors

5-37 *Representation errors* are those that relate to the population unit of interest – sometimes referred to as the 'target population', which, for the census, will usually be persons, households, dwellings or buildings. In assessing the register, the NSO is aiming to understand whether the register will provide relevant information for the target population – including whether or not the register includes relevant units and excludes those not required for the census.

5-38 To assess relevance, the NSO must determine whether or not the set of objects in an administrative data source align with the population of interest for the census (the target population). An 'object' in this context is the basic element of the population on which information is collected, this could be a person, household or dwelling.

5-39 Two key representational indicators are proposed here for establishing relevance of an administrative source for which the following questions are suggested to help guide the assessment:

- *Alignment of the objects in the register with the census target units*
 - What definitions, methods and processes are used to identify and include an object in the source?
 - What laws or regulations define the objects?
 - What checks are carried out by the data holder to ensure that the definitions hold?
 - In the case of misalignment with the census units, is a transformation possible that could meet the census needs?

- *Coverage (of the set of objects) against the census target population*
 - Does the coverage of the objects meet the needs of the census?
 - Is there evidence of under-coverage (objects that are missing from the source - such as undocumented migrants - but are part of the census target population) and/or over-coverage (objects that are in the source, but are not part of the census target population) that would impact the usefulness of the source?

- Are there any differences across geographical areas due to differences in practices by the data holder or due to legislation that need to be considered?
- Are there any rules, legislative or regulatory requirements, including penalties for non-compliance that may impact on the inclusion and exclusion of objects on the source?
- What methods and processes are adopted by the data holder to include (through registration) new objects that meet the required inclusion criteria/definitions and to remove objects (through de-registration) that no longer align with the target population for the administrative source?
- In the case of coverage errors, are there other data sources that could be used in combination with the source to overcome under- or over-coverage in the source?

Measurement errors

5-40 *Measurement errors* refer to the extent to which the particular data items of interest held in the register do not conform to the concepts and definitions of the census topics (such as citizenship, employment status, household size, tenure status, etc.). Thus, the NSO should determine whether or not the information collected about the objects meets the needs of the census, with respect to the target concepts (such as citizenship, employment status, household size, tenure status, etc.).

5-41 More generally is it also useful to know the purposes for which the register holder collects the data, since any policy requirements, legal obligations, targets or incentives (or lack of incentives) could also influence the quality of the data items. Moreover, the possibility of processing error incurred by the data holder in maintaining and updating the register needs to be assessed. Thus, the sort of information that the NSO will need to know in its assessment are:

- whether or not the register includes the variables/characteristics needed for the census;
- whether or not the administrative concepts, definitions and classifications for such variables are consistent with those adopted in the census;
- any differences between the data holder's intended target concepts and what is actually achieved in the register;

- in the case of inconsistency with the census concepts, definitions and classifications, whether or not a transformation of the variables/ characteristics possible to satisfy the census needs;
- the frequency of collection for a variable/characteristic and whether or not each relates to the relevant reference period (and whether or not an indication of the period to which a data item relates is recorded in the register);
- any known delays between a unit (or event) being captured in the register (such as registration of a birth being up to several weeks after the birth itself);
- what incentives (such as benefits) are there for a data subject to provide complete and accurate information to the data holder or update their information, or disincentives (such as penalties) for not doing so, as and when their circumstances change;
- whether or not the data holder's collection process raises any concerns about the quality of the variables, such as the potential for any bias or misreporting arising from the data being recorded by proxy;
- the procedures that are in place to validate and check data on entry by the data holder;
- whether or not data are edited or imputed, and if so, when and how, and if there is an indicator to identify when such an edit and imputation has taken place;
- if there any rules, regulations or incentives on the data holder that may impact on the way data are processed; and
- whether or not the processing carried out by the data holder suggest the quality of the resulting data will meet census needs, and what checks are carried out by the data holder to assure quality.

5.3.1.2 Timeliness and punctuality

5-42 Even if an administrative register may cover the relevant time period for the census, to be useable the data must also be available and delivered to the NSO on time within the census schedule. It is important that the difference between the reference date to which the data refer and the date on which they are supplied to the NSO is kept to a minimum, since the longer the delay the less relevant those data become, even though they may still be accurate. Moreover, punctuality – the difference between the expected date of delivery and the actual date of delivery – is also an issue since the NSO will usually have a

responsibility for producing its census output to an agreed schedule and would not want any delay in the production of the census data to affect this.

5-43 A number of key factors should therefore be considered when assessing timeliness and punctuality of the data source:

- *Time lag at data source:* Besides aiming for a prompt delivery, the frequency of record maintenance and updating of the data by the source keeper can be a key factor in achieving timeliness. Another element that comes into play is the obligation and incentives of the individual to register or deregister with the authorities (such as where there is a change of usual residence).
- *Alignment of data availability with census requirements:* When considering administrative data-based censuses, if the timeliness of supply does not align with the required publication date then there will be inherent issues in the data quality
- *Tentative date of acquisition:* The identification of the dates on which the data from the administrative sources may be available will depend on the processing and format of the data that the administration needs to provide.

5-44 As indicators to assess timeliness and punctuality of the data source the NSO will therefore need to know:

- what the time lag is between an event or phenomenon occurring and being captured in the register (as also noted under 'Relevance');
- how the register's reference period compares with the census reference date;
- whether or not the register has been completely updated when provided to the NSO - and if there is a lag for more complex cases, which might introduce bias;
- what the time lag will be between the end of the register's reference period and the date that it can be made available to the NSO;
- how frequently the data can be supplied to the NSO, set against the needs of the census;
- whether or not data formats and structures, and the method of delivery, meet with the NSO's requirements that could, otherwise, impact on the data supplier's timeline; and
- if the NSO can, after delivery, transform and process the data in time to meet the census timetable.

5-45 In instances where the data are unlikely to be available in time, the NSO may wish to establish whether a provisional version of the dataset can be made available ahead of schedule. In such cases, however, the dataset may be incomplete and subject to higher levels of error. There may therefore be a trade-off to consider between the timeliness of the data and accuracy.

5-46 When assessing the dimension of Institutional Environment (discussed at section 5.3.1.5 below), it is important to include the delivery dates against the data reference periods within any formal agreements with the data supplier. Although the data may be available on time to meet the requirements of the register holder, they may not necessarily be delivered to the NSO in time to meet the requirements of the census – the timely delivery of which the NSO must take sole responsibility.

5.3.1.3 Coherence and comparability

5-47 It is important to assess the degree to which an administrative source can be successfully linked and combined with other data sources for use in the census. The information in assessing relevance - including information about the differences in the underlying concepts, definitions, classifications and collection methods - can also be used to assess coherence.

5-48 As has been noted above, data held in administrative registers are subject to changes over time and across domains due to changes in policy, legislation, and procedures. Such changes can affect the concepts, definitions, classifications and coverage of a source, and can impact on all of the representation and measurement indicators outlined under the dimension of relevance. This is of particular importance for the census, where stability over time is a key concern.

5-49 In assessing the extent of any non-conformity the NSO can use, as indicators the NSO will therefore need to know what changes there have been in the register over time or what differences across domains (such as geographical areas or sub-populations) that may affect:

- the definition and coverage of a unit that is relevant to the census;
- the concepts, definitions and classifications associated with the variables in the administrative source that are relevant to census; and
- the data collection, processing and quality assurance procedures that could impact on the quality of register for census purposes.

5-50 A further key consideration is assessing the coherence and comparability dimension of the data source is the ease with which the data in any one register can be linked with other datasets that are relevant to the census through, for example, the use common unique IDs or, if not, a combination of other identifying variables such as name, sex age and address of each data subject. (This element of *linkability* is particularly important in the assessment of the quality of each characteristic to be used as a census variable and will be discussed further at the input quality stage in the next section.)

5-51 As indicators of linkability the NSO will want to identify:

- The presence of a common unique ID key, by asking:
 - Does the source include a unique identifier that is common with the unique keys required for the census linkage?
 - If so, is the identifier available for all of the relevant objects on the source? And if not:
- The presence of a common unique combination of variables by asking:
 - Does the source include a unique combination of variables (such as name, age and address), which could be used for the census linkage?
 - If so, are the unique combination of variables present for every object on the source?

5.3.1.4 Accessibility and interpretability

5-52 A number of indicators can be developed for the assessment of the quality as determined by this dimension

What restrictions are there on access to and use of the data source?

5-53 It is important to identify any restrictions that may impact on the NSO's ability to access and use an administrative source. For example, existing data protection restrictions embedded in legal acts can impose certain limitations on the data acquisition and processing. Such legal acts may be specific to particular data sources or may be more generic allowing the NSO access to such data sources as and when required, subject to the agreement of the register holder.

5-54 However, the register holders themselves may also impose further restrictions on the supply of data and the permitted use. These can include:

- suppression of records or variables;
- disclosure treatments (pre-delivery), such as encryption of identifiers, perturbation, banding or top-coding of the supplied data;
- restrictions on how the data can be used;
- restrictions on the retention of data and rules on deletion/disposal of the data;
- rules on disclosure methods that must be applied by the NSO, affecting the census outputs.

5-55 The NSO should establish and describe any restrictions that apply, on which an assessment can be made as to the impact (and risks) of the restrictions on the use of an administrative source in the census. As part of the assessment, the NSO should also consider whether or not it has the capability of applying the necessary technical and procedural safeguards to comply with any such restrictions abide by the restrictions. The safeguards often form part of a Memorandum of Understanding (MoU) or Data Security Agreement with the data holder.

What is the level of public acceptability?

5-56 Whether or not an NSO decides to access a particular data source for use in the census may also depend on public acceptance. The NSO should therefore be transparent about the use of administrative sources in the census, highlighting the benefits to the public, whilst providing assurance against privacy and security. To assess public acceptability, the following tools or processes can be used:

- A *public consultation or engagement programme* may be carried out by the NSO on the use of administrative data in the census (or for other statistical research or outputs). This can take various forms, including formal consultations, questionnaires (through surveys or via the NSO inviting feedback via its website), and/or qualitative research into public attitudes by bringing together members of the public - selected to be representative of the population, or to reflect different population groups of interest - to assess their views and opinions.
- As a risk management tool, used in the planning phase of the census, a *Privacy Impact Assessment* may be carried out to assist organizations to consider more fully the privacy implications of the way that the data is collected and processed.

- Alternatively (or additionally) a *data ethics assessment* might be carried out to establish whether the access to, and the use and sharing of public data, for research and statistical purposes, is ethical and desirable for the public good. For this purpose, NSOs may choose to employ a formal and independent body to provide expert advice or endorsement, such as a National Data Ethics Advisory Committee (if such bodies exist within the country).

How easy is it to transfer data?

5-57 The data supplier might adopt very different data models, formats, schemas, software and hardware to that with which the NSO is familiar. This includes how data are held and transmitted, (including the security arrangements for transmission). The data structures could also be complex and file sizes extremely large (particularly for transaction data). It is important that the NSO understands such differences and complexities, in order to assess whether or not it is feasible to receive and transpose the datasets into the NSO's systems. This process can also include negotiations with the supplier on the development of processes and systems to facilitate the transmission of datasets in a format that meets the needs of the NSO. However, this can be a time consuming and costly process.

5-58 More generally, cost is a key factor to be considered when assessing ease of access. This can include costs imposed by the data supplier, or costs incurred by the NSO in developing its capability to receive an administrative dataset (for example, if new software or hardware needs to be purchased). It is important to assess any costs against the expected value a new administrative source will bring.

5-59 In practice, details of the arrangements for the transmission of data to the NSO (including the files structures and format, the variables, the frequency of supply and dates for delivery, data standards and agreed costs) should be included in appropriate Data Sharing or Delivery Agreements between the NSO and the supplier.

Is there clear and comprehensive metadata?

5-60 An assessment of interpretability relates to the existence and availability of comprehensive and clear metadata and documentation about the

administrative source. Without this, it is not possible to understand and assess the administrative source against the intended use. The metadata should include details about the:

- administrative organization and the purpose of the collection
- concepts, definitions, classifications and protocols used
- collection, processing, validation and quality assurance methods and procedures
- reporting units and variables; including data dictionaries, file structures, formats and relationships within the data

5-61 This information is also important for the assessment against the other quality dimensions discussed in this chapter. It will often be the case that clear and complete metadata will not exist for all aspect of an administrative source in the initial phase of exploring the source for use by the NSO. It is therefore necessary to work with the data supplier to build the relevant metadata. This relies on good communication with the supplier and a willingness of the supplier to work with the NSO (see section 5.3.1.5 below). Depending on the complexity of an administrative source, an NSO may decide to appoint staff to work within the administrative organization to develop an in-depth understanding of a source.

5.3.1.5 Institutional environment

5-62 In moving to a register-based methodology the NSO becomes reliant on the data holders to collect, process, and deliver the administrative data to the quality expected and to the agreed timetable. It will be beneficial for the NSO to determine rules and processes to ensure interoperability not only between information systems in public administrations, but also to facilitate access to administrative data for statistical purposes.

5-63 Such processes will include the development of appropriate infrastructures or mechanisms, such as the compilation of inventories of registers that cover the relevant administrative data collections across the whole realm of public administration - as has been developed in Hungary and Estonia for example. In Estonia, in particular, changes in the administrative databases are recorded centrally and are communicated to the NSI by the relevant authority. In Norway, a central hub has been created in which all the administrative data held by public authorities are deposited, and where no information that is already held by the hub may be collected again a public authority directly from citizens, households or enterprises. In Cyprus, a central

government data warehouse exists where most of the administrative information is collected and maintained, and which is accessible by the NSI and all public authorities. In the France there are policies for publishing government data in open formats, while in Poland and Moldova are developing the technology for standardised systems that would enable interoperability among government databases.

5-64 The NSO is also reliant on the quality of the metadata that the data supplier provides about the dataset (as noted at sub-section 5.3.1.4 above) and about any expected changes to the data. It is therefore important to assess confidence in the data supplier's ability to meet these needs.

5-65 Therefore, in order to assess the ability of the register owner to provide the data necessary for the purpose of the census, the NSO should develop a good working relationship with the supplier with the aim of:

- ensuring that the NSO's requirements and quality standards are known to the supplier;
- establishing whether or not any issues have arisen regarding the quality of previous supplies of data;
- determining if there any changes or proposed changed (relating to, for example, the legal basis, concepts, classifications, or the processes and procedures for the collection, management, supply and protection of the data) that might affect the quality of the data source; and
- feeding back findings to the supplier that could result in improvements to the source.

A number of indicators are therefore suggested for the assessment of the institutional environment:

The strength of the inter-relationship.

5-66 There should be processes in place for managing the relationship with the data supplier; ensuring there is a continuous dialogue. These should include mechanisms for:

- the communication of the NSO's requirements to the supplier
- the timely communication (by the supplier) of any changes that might affect the data source (such as changes to the legal basis for the data, to concepts and classifications and to the processes and procedures for data collection, management and supply);
- raising any questions with the supplier about the data source; and

- feeding back findings to the supplier that could result in improvements to the source.

Previous experiences

5-67 The NSO should assess how responsive a supplier has been to any previous requests for data, whether or not any issues (such as late delivery, poor data quality) have arisen with any such previous supplies of data

Formal agreements.

5-68 The NSO should seek to establish or develop written agreements (legal or otherwise) exist or can be developed, covering:

- roles and responsibilities of the NSO and supplier, including whether or not the NSO has a role in the approval of any changes to an administrative source used (or to be used) in the census
- the legal basis for the supply of data and any security/confidentiality requirements
- the specification of requirements, as set out in any Sharing/Delivery Agreement noted at sub-section (3) above.

The status of the supplier

5-69 The risk associated with the status of a supplier should be assessed by the NSO, considering whether or not the supplier is an established, stable and reputable organization. This should consider whether there is any legal or regulatory basis to the administrative function the supplier carries out that would give confidence in the sustainability and quality of the source. Risks associated with the complexity of the supplier organization(s) involved in the collection, processing and delivery of the source should also be considered. For example, there could be multiple bodies or organizations involved, each impacting on the quality of the final data supply.

Supplier's quality standards.

5-70 An assessment of whether or not the supplier can meet the quality expectations of the NSO should be made. This should consider the information on the principles, standards and guidelines adopted by the supplier for assuring quality, including the procedure in place covering collection, processing and the supply of data to the NSO. Evidence of how the supplier checks whether the

standards are being met (for example, whether through internal or external audits by regulators or professional bodies) could be valuable. The supplier may also produce quality reports, which should be reviewed by the NSO. A more detailed assessment based on key aspects of the administrative source is included under the relevance quality dimension above.

5.3.2 Input quality

5-71 The quality of the register data to be used in the census can be affected by several factors. In addition to those that may influence the quality of the register as a source in general, other factors will influence the quality of the supplied data units and variables. The assessments of these are summarised in this section with respect to the relevant quality dimension in each case.

5.3.2.1 Accuracy and reliability

5-72 As noted at sub-section 5.3.1.1 above, in assessing the accuracy of the input data NSOs should distinguish between ‘representative errors’ (those relating to the coverage of target population – as noted in the previous section) and ‘measurement errors’ (those relating to the particular variable being considered).

5-73 *Representative errors* might occur if units are not reported correctly or promptly to the administrative authority (as a result, for example, from delayed entry in, or removal from, the register) or if some data records are not transmitted to the NSO because of technical problems. The extent of such errors can be assessed by estimating the level of over-coverage (duplicate entries) or under-coverage (missing entries) by comparison with other data sources. For example, comparing the population and social security registers may reveal that deceased persons are still being reported by a country’s population register though identified as deceased in a social security register. And school-aged children in the statistical population register should be largely covered by the register of enrolled pupils and students.

5-74 To assess *measurement errors* the NSO will need to be able to estimate, for any particular variable, the number and percentage of both missing values and those values that are out of range or invalid (such as an age recorded as 120). Statistical techniques and metrics, such as frequency distributions, can

reveal unexpected patterns and outliers which might indicate systematic measurement errors.

5-74 A number of indicators can be used to measure the accuracy of the supplied units for an assessment of any representation and measurement errors in the data (as described at section 5.3.2.1 above).

Assessment of representative errors

5-75 Basic indicators to assessment representation errors include:

- the total number of units received (for comparison against expected count);
- the percentage of duplicate units.

5-76 A broad assessment of under- and over-coverage of the data can be made by computing and comparing the total number of units, as well as cross-tabulations of frequency/percentages across key characteristics (such as sex, age, geography) on an aggregate level, between the administrative source and other/alternative sources taken as reference or a comparative 'gold standard'.

5-77 A key indicator in assess under-coverage in this way would be:

- the percentage of units in the reference source (the register) that are missing in the supplied (administrative) source.

while over-coverage can be assessed by:

- the percentage of units in the source that are not included in the register population; and/or
- the percentage of units in the source not belonging to the target resident population of the NSO.

5-78 This approach is subject to two assumptions. Firstly, a complete base register (the register that is assumed to have the most relevant coverage of the target resident population, including the target population of the input data) must be available to compute over-coverage. For instance, immigrants may not be included in a country's central population register but may be identified as new residents in resident permit records. Secondly, it should be clear which units of the complete base register should be included in the input data to compute under-coverage. For instance, school-aged children in the statistical population register should be largely covered by the register of enrolled pupils and students.

5-79 Another approach in assessing under- and over-coverage is to conduct a special sample survey that provides two possible ways to make comparisons: (a) conducting an independent coverage survey to measure under-coverage; and (b) checking a sample of the supplied source in the field to measure over-coverage.

5-80 Finally, comparisons can be made at the unit record level to determine the percentage of units that are consistent within and across sources, described elsewhere as a measure of their 'authenticity'. An example of inconsistent units might be where each unit or row within an administrative dataset represents a registration event (in health records for example) that includes, name, address code, PIN, together with date of registration. Two records relating to a single person are inconsistent if the period of registration of the person at different addresses overlaps. The percentage of inconsistent units provides an indicator of error or conversely, accuracy. However, unit-level analysis has its limitations as sources may differ at the micro level but may nevertheless result in similar statistical measures such as means, medians, etc. In addition, where unit-level comparisons are made between multiple sources, it is important to note the possible impact of selectivity bias within the linkage process on any resulting differences.

Assessment of measurement errors

5-81 Basic indicators to measure the completeness of the characteristic variables supplied within administrative datasets at the aggregate level (such as age, sex, ethnicity, etc) include following:

- number and percentage of missing values within key variables (such as date of birth and sex);
- number and percentage of out-of-range values within key variables (for example a recorded age of 120 years);
- number and percentage of implausible values (based on, for example, cross-tabulations of different variables);
- prevalence of unexpected frequencies, patterns or outliers, based on frequency/distributional analysis of key variables (aggregate comparisons with external sources, as well as expert knowledge, can also be used to identify data oddities);
- the degree of consistency of the supplied data at the aggregate level - namely that relationships between related variables are consistent and not implausible provide a measure of the accuracy of variables.

5-82 Statistical techniques and metrics, such as frequency distributions, can reveal unexpected patterns and outliers. For example, a cross tabulation of age and marital status may lead to the identification of implausible cases, such as a 5-year old child that is recorded as married.

5-83 Similarly to the assessment of representation error, an efficient way to assess variable accuracy, especially in the preliminary analysis of data and/or the very first time the data are used, is the comparison of data; that is, where the input data are checked by means of comparison with other independent sources that contain the same variable. Suitable independent sources for comparison could include a national survey (such as a labour force survey) or a previous census.

5.3.2.2 Timeliness and punctuality

5-84 It is important that the difference between the reference date to which the data refer and the date on which they are supplied to the NSO is kept to a minimum, since the longer the delay, the less relevant those data become, even though they may still be accurate. This gap between reference date and acquisition by the NSO is referred to as *timeliness*. However, *punctuality* – the difference between the expected date of delivery and the actual date of delivery – is also important as the NSO will usually have a responsibility for producing census outputs to an agreed schedule and would not want any delay in the supply of the census data to affect this.

5-85 Measures of *timeliness* can be determined relatively easily by comparing the reference date, the specified delivery date, and the actual delivery date of the data. Thus, two aspects of timeliness can be assessed by:

- the difference between the date of capturing any change to the data in the source by the register holder and the date on which the change actually occurred in the population; and
- the difference between the date of receipt of the data by the NSO and the date of the reference period to which the data refers,

whereas *punctuality* can be simply assessed by any delay suffered in the receipt of the data by the NSO from that originally agreed with the register holder.

5.3.3 Process quality

5-86 As data held in any administrative source will not have been collected for statistical purposes, they must be transformed by the NSO in some way for use in the census. Such transformation processes will usually include (as discussed in Chapter 4):

- linkage to the data across different administrative sources;
- integration into a statistical register;
- duplicates and conflict resolution;
- updating registers using a ‘signs of life’ approach
- cleaning and editing (to rectify errors in the data); and
- imputation (to resolve missing values and inconsistencies).

5-87 Each of these processes is informed by the information gained through the source and input quality stages. For example, if a data source is found to have erroneous or missing values at the input quality stage it may be necessary to edit or impute the data for use in the census.

5-88 These processes are, of course, designed to improve the quality of data and to make the output products fit for purpose. However, data processing in itself may introduce a degree of uncertainty in the data (either inadvertently or, in the case of statistical disclosure control, deliberately) that, if not controlled, can reduce the quality of the output. The change in quality due to the implementation of such processes is referred to as ‘process quality’. This section focusses on some of the key considerations around these processes that may affect the accuracy and completeness of the data.

5.3.3.1 Accuracy and record linkage

5-89 As noted in the discussion on source quality above, if the NSO relies on more than one source of administrative data for its census dataset, it is necessary to be able to link the unit record data from the different sources. The degree of *likability* will affect the accuracy of the input data. The quality of the linkage of unit record data from different sources is (as noted in Chapter 4) dependent on the availability, uniqueness and prevalence of common identifiers.

5-90 For countries that adopt a unique identifier that is common across their administrative registers the process is relatively straight forward. Such an

identifier reduces the effort required to interlink the data by, in particular, easing the process of evaluating completeness and accuracy. To assess this dimension of coherence in the data from different sources the NSO will need to determine the number and percentage of records supplied where such a key is missing and, in such cases, ensure there is no systematic error.

5-91 But where such an identifier does not exist, the quality of the linkage can be a major concern. In such cases, record linkage using match keys (that is, variables that are common to the units in each data source - typically, name, date of birth, sex, and address) may be possible but this is usually more time consuming. In this case, the NSO needs to be assured that such 'linking' variables have excellent quality in all sources, otherwise the quality of the subsequent linkage, and thus the reliability of the data, will suffer. The risk of false matches (and, conversely, false non-matches) should be assessed. (For more detail see the *ESS Handbook for Quality Reports*³⁰.)

5-92 The match keys may be linked using *deterministic* linkage (where the same set of match-keys exist on each dataset being linked) or by *probabilistic* methods (where exact matches are not possible and where links are made based, instead, on comparing and qualifying the *relative similarity* of records across the identifying variables). It is important that these variables are accurate and complete to ensure the high quality of record linkage. The accuracy of linkage can be assessed through differences between the characteristics of linked and unlinked records.

5-93 One additional linkage method that can be applied to unlinked records after deterministic and probabilistic methods have been applied, is *clerical linkage*, which involves a visual inspection of the unlinked records, but this can be laborious and time consuming, and, moreover, is not possible when the data are 'hashed', that is when an algorithm is applied to each data item in the original data to create a code that uniquely identifies the information but masks the identity of the original data.

5-94 Linkage error can occur through unlinked records that should have been linked (also known as 'false negatives') and linked records that should not have been linked (also known as 'false positives').

5-95 Two common methods for assessing linkage quality are:

- Estimation of false positive and false negative rates, through for example a clerical review of a sample of linked records, although it should be noted that such a review can only be done when the data are not hashed. If the data are hashed, the NSO should try and obtain access to a sample of the linked records with all the original information on the data to assess the linkage.
- Comparison of the distributions of characteristics of linked and unlinked records, by for example such variables as age, sex and ethnicity. Differences in characteristics suggest that some bias is introduced by linkage error, resulting from the fact that certain types of unit records) will not be linked because they are more difficult to link.

5-96 Regardless of whether a unique key or identifier variable is available or whether several variables will be used in combination to identify matches in the linkage process, there should be indicators to assess the quality of linkage variables supplied, such as:

- the number and percentage of duplicate linkage keys, which can be calculated for a single identification variable (such as a personal identification number), or a combination of variables which provide the linkage key (such as age, date of birth, address); and
- accuracy indicators as described in previous sections (including the absence of linkage keys, or the number of missing or implausible values) which may show the extent to which measurement errors are not random – for example where there is evidence of any systematic biases (for example where certain population groups may be subject to greater error against the linkage key variable(s)). Therefore, it is important to calculate these indicators by geographic areas and population sub-groups.

5-97 Finally, if and where the linkage variables have been provided to the NSO in an encrypted or ‘hashed’ form (that is masked using a one-way algorithm to protect the privacy of the data subject), it must be verified that the hashing performed by the supplier matches the hashing algorithm used at the NSO, otherwise, it will not be possible to link the data supplied to other data sources, undermining the relevance of the data.

5-98 But a word of caution here - the more registers that are linked together in this way for the purpose the census, the higher the possibility that the timing of their updates may be a risk factor in ensuring quality of the information – particularly in respect of the dimension of simultaneity. NSO should therefore be kept informed of the timing and frequency of updates for all the relevant registers to ensure synchronization of the input data.

5.3.3.2 Assessing over-coverage in statistical population registers by using signs of life

5-99 Coverage issues can arise in any administrative source (as noted in the section on input quality) and coverage checks are necessary before data integration can take place. In addition, coverage issues can also arise following the integration of administrative sources when constructing statistical registers (see section 4.3). Within the process quality stage, NSO should pay particular attention to assessing (and reducing) over-coverage as this greatly impacts on the dimension of accuracy.

5-100 An increasingly commonly used tool to help minimise over-coverage is the so-called ‘signs of life’ (SOL) method in which a number of ‘activity rules’ are adopted to ensure that only persons who are alive and meet a set of pre-defined residency criteria are included in the census. The data from those administrative data sources to which the NSO has access can be cross-checked to improve the coverage of statistics derived from a population register by using the information from the different registers that relate to the activity (signs of life) of the same resident at a particular period in time (see section 4.4 for more details of the application of the SOL methodology).

5-101 The success of a SOL method relies on the availability of good indicators of signs of activity in the individual or combined administrative data. The application of the method typically involves making some assumptions, which determine who is considered as active and who is not. The choice of signs of activity indicators (or decision rules) should be informed by an assessment of quality at the source and input data stage (sections 5.3.1 and 5.3.2 respectively), including consultation with data suppliers, cross-validation between sources and over-time, and expert opinion.

5.3.3.3 Assessing coverage in statistical population registers by using independent surveys

5-102 In addition to the signs of life approach there are several other methods are available to estimate the coverage (and, indeed, content error) of censuses. These include: simple quality assurance techniques such as internal consistency checks; demographic analyses; comparisons with data from other sources including previous censuses and/or current household surveys; and post-enumeration surveys. Such methods will be appropriate regardless of the census methodology adopted, but it is perhaps worth stressing the important role that a post-enumeration survey can play in the assessment of the quality of the statistical register.

5-103 The post-enumeration survey (PES) is a complete re-enumeration of a representative sample of the census population followed by matching each individual enumerated in the PES with the data collected from the census (whether that is through a field enumeration or derived from registers). Using dual system estimation (DSE) processes, the results of the comparison are mainly used to measure coverage (and content) error in the census. It is desirable to design the survey so that reliable estimates of under-count or over-count are obtained for sub-national geographic areas as well as for the country as a whole. In addition, some estimates can be made through post-stratification by forming subpopulation groups specified by sex and age group. The survey may also be used to measure levels of agreement for the data collected on selected characteristics, such as sex, age, marital status, or relationship to the reference person or head of household.

5-104 It is important, however, that, in order to avoid the same systematic errors that might be present in the census data itself, the PES should be conducted independently from the management of the statistical register.

5-105 It is not intended that the remit of this handbook should discuss in any detail the indicators that will demonstrate how, generally, the PES and DSE methodology can be used to quantify the level of under- or over-coverage in the census. Instead, readers are invited to refer to the UN's *Post Enumeration Surveys Operational Guidelines Technical Report*²⁶. It will be sufficient perhaps just to note here that the basic stages to measure and adjust for the number of people and households not counted in the census or counted in the wrong place as recorded in the statistical register generally comprise:

- matching the PES records were with those from recorded on the statistical using a combination of automated and clerical matching;
- using the matched records and the DSE technique to estimate the number of people and households missed by both the census and the PES;
- searching for duplicates in the register and in the PES to estimate the level of overcount (those counted more than once);
- estimating the populations for each local area by age and sex by balancing over- and under-estimates, using a combination of statistical regression and small area estimation techniques; and (if necessary)
- imputing the missing people and household into the census database/statistical register.

5.3.3.4 Methods for deciding on variables reported in different sources and conflict solutions

5-106 When using data from multiple sources, methods are required to assess the quality of – and hence inform the choice of - variables where the same attribute is reported in different registers. For example, the Spanish *Padrón* does not record a person’s legal marital status. So, in order to assign this attribute data from several different registers are used. But because an individual may appear in multiple data sources with conflicting information, decisions rules are necessary (as with the SOL method described above) to determine which value is accurate.

5-107 As noted at section 4.3.3, there may also be a need to resolve conflicting descriptions of addresses recorded in different registers and to identify the most accurate information. Such differences may arise, for example, from delays in individuals reporting a change of address or reporting the address of a second home. Conflicting (or multiple) address information and any related decision may cause under-coverage in some areas and over-coverage in others. The NSO will therefore need to decide to which address the census information should refer (perhaps by using, for example, the most recently reported address).

5.3.3.5 Assessing the quality of the editing and imputation process

5-108 Editing and imputation is an iterative process of reviewing the data in order to correct any errors resulting from invalid, inconsistent or implausible recorded entries and the subsequent substitution of these, together with any missing data items, with plausible values, thereby improving the accuracy and

coherence of the data and making them suitable for the purpose for which they were collected.

5-109 In this process, the data is first edited - in other words, checked - for missing, invalid or inconsistent values, and then imputed for resolving such problems, and finally validated for ensuring the quality of imputation. Data editing must be repeated after the data are imputed, and then again if the data are subsequently further altered during any process for protecting individual information (statistical disclosure control)⁹.

5-110 Editing and imputation may be required both on the single source and the integrated data. As has been previously noted, base registers are used mainly to provide information on the population count and basic information on the respective census topics. Other specialised or supplementary registers are used for providing the information on particular census variables, and sometimes for providing data that are either fully or partly missing in the base registers. Such specialised registers are also used for validation purposes by which the dataset that has been combined from the base and other registers, (referred to as the 'integrated statistical register') is enhanced with imputations for item non-response and implausible values, thereby creating the final dataset from which the final census outputs are produced.

5-111 The quality of the editing and imputation processes should be systematically monitored and evaluated. It is suggested that the original value, edited value, type of error, and data on failed checks should be monitored for analysis of editing changes. For every cycle of editing and imputation, change analysis should be undertaken based on the following types of information:

- i) frequency and type of edit flags;
- ii) magnitude of change; and
- iii) extent of redundancy in consistency checks.

5-112 For monitoring the application of editing rules and imputation procedures, a number of indicators can be used. The following indicators are suggested:

- (a) Edit failure rate: This is the number of detected errors divided by total errors (detected and undetected errors) in the input file. This indicator varies between 1 and 0. High value means the editing rule is properly defined as a criterion for error detection. If it is low, then the rule fails and must be improved.

(b) Adjustment rate: This is the number of households or people created under absent household imputation (if used) divided by total the number of households or people.

(c) Imputation rate: This is the percentage of imputed records in the total number of records. This indicator should be calculated for each variable. Imputation rates are key quality indicators for the census as a whole, and for data users. Unit imputation must follow a statistically robust, transparent process, and be well documented.

(d) Dissimilarity index: This represents the degree of change of two distributions (observed and total including imputed values) at the variable level. This index, calculated as:

$$ID = \frac{1}{2} \sum_{k=1}^K |f_{y_k} - f_{y_k}^*|$$

(where: k = categories of the variable: f= percentage distribution of the variable before imputation; f*= percentage distribution of the variable after imputation) ranges from 0 to 100, where 100 implies maximum dissimilarity.

5.3.4 Output quality

5-113 Regardless of the type of data collection methodology used for the census - traditional, register-based or a combined approach - it is equally important for NSOs to assess the quality of the output so as to ensure that the census produces high quality data that are fit for purpose. The methods of doing so will be broadly similar whatever the type of census. However, it will be helpful in the context of this handbook to consider the issues of output quality within the context of a register-based system since it is particularly important for those NSOs that have moved to a register-based or combined census data collection methodology to assess these issues in order to determine whether or not the transition has affected the overall quality of the outputs.

5-114 Assessing output quality should not be restricted to the estimation of overall uncertainty of the estimate (the accuracy dimension) but should also include an assessment across all other quality dimensions. The introduction of administrative data will likely lead to initial gains in some dimensions and losses in others. Achieving the right balance across the quality dimensions is therefore the key to best meeting user needs.

5-115 It is therefore necessary not only to consult users throughout the census design process, but also to give them access to the general information and specific metadata they need to understand the quality of the data. The opportunity to then provide feedback on the NSO's quality assessments will assist the processing of improvement in output quality in subsequent censuses.

5.3.4.1 Relevance

5-116 As defined at paragraph 5-8, 'relevance' refers to the degree to which the census outputs meet the needs of users in terms of both coverage and content. This dimension may require NSOs to adjust the direction of their output programmes as user requirements develop over time. However, assessing relevance is subjective because it often depends on varying (if not conflicting) user needs. The challenge, therefore, for a census output programme is to balance any conflicting user requirements and to go as far as possible towards meeting the most important needs within resource and other constraints.

5-117 The three most common approaches to assessing relevance are:

- to ask users directly either through a user needs survey (before the census) or a user satisfaction survey (after the census);
- to instigate ongoing stakeholder feedback mechanisms; and/or
- by analysing the uses made of the data in a census use study by, for example, reviewing published reports and papers and/or downloads of data from the NSO's website and requests for commissioned outputs.

5-118 The level of compliance with international recommendations (or obligations to supply statistical data to fulfil international commitments) can provide some measures of relevance at the global level.

5.3.4.2 Accuracy

5-119 Measuring the overall accuracy of final census counts has traditionally been based on the dual system estimation (DSE) which (as noted at paragraph 5-16) usually involves carrying out an independent large sample post-census coverage survey (Post-Enumeration Survey) and then adopting capture-recapture methods to estimate under- and over-coverage. The resulting estimates can then be adjusted based on administrative data on births, deaths, and migration flows, for each year between decennial censuses. Additionally, some NSOs have traditionally carried out smaller sample post-census quality

check surveys where data is collected on all the census questions and then matched to statistical dataset, in order to measure respondent error.

5-120 In a register-based census methodology, the traditional methods for determining overall coverage and quality are, to a certain extent, still applicable. However, in the case of census counts produced solely from administrative registers, new or revised methods are necessary. In particular, the traditional DSE method does not deal well with over-coverage, as it introduces bias. Moreover, after the several processes to assess quality in the source, input and process stages, some NSOs may adopt of strategy that demands less (if any) reliance on the DSE approach to assess coverage – relying more on the quality of the maintenance and updating of the population registers to do so. Many countries, for example, have attempted to develop methods which aim to reduce virtually all over-coverage through ‘signs of life’ methods noted in the previous section.

5-121 Though such new and revised methods are of increasingly significant interest among NSOs, with ongoing developments across a number of countries (see *The Survey Statistician*³⁶ for a summary of new and emerging methods) It may, nevertheless, still be advantageous to adopt traditional DSE methods as well.

5-122 However, one further measurable process - applied to the census outputs prior to their public release - that may also affect the accuracy of the data are the measures adopted to ensure confidentiality by protecting against statistical disclosure. Ensuring data quality and protecting the confidentiality of personal information are in some sense conflicting aims, but both are essential aspects of census output. It is important to understand that – regardless of methods of data collection - it is difficult, if not impossible, to publish completely accurate outputs (especially for small areas) while, at the same time, adopting sufficient levels of disclosure control to ensure the protection of confidentiality. (For more information about statistical disclosure control methodology see the ESS Handbook on Statistical Disclosure Control³¹.)

5.3.4.3 Timeliness and punctuality

5-123 As noted at section 5.3.2, ‘*timeliness*’ refers to the lapse of time between the date to which the census data refer and the date of publication of the data. A combined or register-based census offers the potential for census outputs to be produced in a more timely and frequent manner than in a traditional

decennial census – indeed, this is one of the principle advantages of the transformation from a traditional to a register-based system.

5-124 In light of this, the timeliness of estimates that can actually be produced should be a key quality consideration. Of course, the timeliness of the outputs is to a large extent determined by the timeliness of the delivery of the input data (as discussed above) as well (of course) as the time taken by the NSO to do all the data processing. However, there is often a trade-off between timeliness and accuracy since, generally, the more time allocated by the NSO for ensuring the accuracy of the data during processing the longer it takes to deliver the outputs to the user. In some cases, it may be preferable for the NSO to wait to get higher-quality input data, or a second data supply for the same reference date. Moreover, It may be the case that different users will have different views on the balance between the two, and as such they may not have the same view on the effect of improved timeliness vis-à-vis accuracy.

5-125 Quantitative indicators can be applied to measure the timeliness of the final results including any delays between data collection, data acquisition, data linkage and publication of statistics. For example, the overall timeliness may be calculated as the time from the end of reference period to receiving administrative data supply, divided by the time from the end of reference period to publication date, multiplied by 100 per cent^{2,30, 37}.

5-126 The related dimension of ‘punctuality’ can be more easily assessed by simply determining the difference between the planned date of delivery of the outputs and the date that such delivery was actually achieved.

5.3.4.4 Coherence and comparability

5-127 Measuring the extent to which census counts produced using administrative data are internally and externally coherent and comparable is an important aspect of output quality. Such estimates should be coherent with the known characteristics of the population both longitudinally (by comparing one census with another) and geographically (by comparing one area with another) as well as being internally consistent with the outputs from the NSO’s other statistical data operations. It is also important to assess the extent to which output data are comparable to, and consistent with, international standards.

5-128 Demographic analysis (DA) is one particular approach that can be applied to assess the coherence and comparability (as well as the accuracy) of census

outputs. DA involves systematic comparisons, establishing thresholds of acceptability and understanding any significant discrepancies. The census data that integrate administrative data are validated against alternative sources – such as survey data, previous census data or alternative sources. When using DA, it is important to keep in mind that estimates in two sources can be different across different sex-age or other distribution. These differences could be caused by:

- different target populations,
- different reference dates or population changes (particularly when comparing to historical census data),
- conceptual differences and variations in classification between the variables being compared across different sources, and/or
- by differences in sampling, collection methods and approaches to data processing.

As such, any such comparisons must be made in light of the results of the assessment at the source and input data stages.

5.3.4.5 Accessibility and clarity

5-129 Within the context of the dissemination of outputs, accessibility refers to both the ease with which the existence of the data can be ascertained by the user, and the suitability of the means and formats by which the data can be made available to the user. Clarity – or interpretability - relates more to the availability of any supplementary information or metadata that may be necessary of help the user to interpret and understand the statistical outputs.

5-130 These dimensions can be readily assessed by comparing the means of dissemination with the initial set users' requirements determined through consultation beforehand. It is often helpful for the NSO to develop a user satisfaction survey and/or convene an expert review panel once the output programme has been completed to assess the degree in which accessibility and clarity (as well as relevance as noted at section 5.3.4.1 above) have been met from the users' perspective.

5.4 Reporting quality

5-131 Whether or not administrative data are used in statistical production, assessing the overall quality of estimates produced should take into account each of the quality dimensions discussed in this chapter. This includes not only the accuracy dimension – the aspect which is most often reported in relation to census methodologies – but also the remaining quality dimensions. In a census context, the overall quality of estimates is thus about establishing the balance across the quality dimensions which best meets the needs of census users. To achieve this, it is necessary to consult users throughout the census design process, but also to give them access to the general information and specific metadata they need to appraise quality decisions and feedback on quality assessments undertaken by the NSO.

5-132 NSO must always bear in mind that the end users of the statistics must have confidence in the methods used and the quality of both the source data and the statistical outputs. This requires the use of reliable methods that can be readily understood by users. As such, quality reporting and quality metadata are essential. In order to enable the producers and users of statistics to appraise and feedback on quality decisions and determine whether the right balance has been achieved across the quality dimensions, sufficient metadata on the quality assessment is necessary.

5-133 Thus, the final activity of the quality assessment process - covering all the stages from source to outputs - should be for the NSO to publish a report as part of its output programme to document the results of quality assessment and assurance throughout the census production cycle. This report should not only include information relevant to each quality stage but should include recommendations for improvements for future censuses based not only on the NSO's own assessments but the feedback from users and external experts.

6 Conclusions and recommendations

6.1 Summary conclusions

6-1 Administrative registers generated by governments and service providers are becoming an increasingly important source of data for official statistics. In this context, there is a growing interest in the use of administrative sources for producing census information. Administrative data can be used across the different census methodologies (traditional, register-based or the combined approach) and to support all stages of the census process, including constructing an address frame, supporting field operations, enumerating the population, collecting census variables, quality assurance, editing and imputation, and estimation.

6-2 Their use can provide more frequent and timely statistics about the population; improvements in accuracy and reliability; and significant reductions in costs and respondent burden. Many countries that have used traditional census or combined censuses with a large scale of field data collection have faced particular challenges in doing so during the Covid-19 pandemic and have had to postpone their censuses for more than a year. In order to reduce the impact of similar disruptions to the planning of future censuses, there is now more interest in the use of administrative data together with online data collection methods.

6-3 But there are also many challenges to be faced in moving to a register-based statistical system that are related to the quality of existing administrative registers, establishing the necessary legal framework for the use of individual records for statistical productions, and developing the institutional capacity for establishing and maintaining statistical registers. But perhaps, the most significant is that administrative data have, in general, not been collected for the purpose of a census. As such, the NSO may have little control over the concepts and definitions used; the target population; the collection, processing and quality assurance procedures; and the data methods, structures and systems used.

6-4 Moreover, changing the design of the census processes – particularly the data collection operation - from traditional methods to other approaches with the use of administrative data is not easy task and can be more complex and complicated compared to the well-established design of traditional census. In

particular there are a number of necessary pre-conditions that need to be in place before any such transition to a register-based methodology can be attempted among which are:

- the existence of the legal authority for the NSO to receive and use administrative data for statistical purposes, and the requisite political support for the NSO to do so;
- public approval for the use of personal information for such statistical purposes;
- good cooperation between the NSO and the holders of the administrative data; and
- a population register and an established system for accurately linking unit records between (potentially) many different data sources.

6-5 With this last point in mind, it should be noted that for the purposes of the census it may be necessary to integrate data not only from other administrative registers but also from other data sources, such as a field enumeration, on-going or ad hoc sample surveys or a longitudinal study of a census-based sample. Where administrative registers are used as the primary source, these will comprise a combination of 'base registers' (those that hold the basic information relating to the stock of the entire set of 'population' units that are being counted in the census – typically persons, dwellings and buildings), and 'supplementary or specialised registers' from which one or more of the attributes that may be required for the census can be derived.

6-6 More often than not, however, it is neither practicable nor even possible to collect all desired census variables from registers. There will always be circumstances in which it is difficult, impossible, illegal or too costly to obtain some variables from registers, and where NSOs must rely on other data sources either to improve the quality of the data obtained from administrative registers or to supplement such data in order to report on those census characteristics and variables that cannot be derived from the registers themselves.

6-7 The process of linking unit records from different data sources can be done by matching individuals based on name, sex, date of birth, and address. But in practice this is complicated and time-consuming and far from perfect because many people may share some of the same characteristics and/or because some of the data may be missing, incomplete or erroneous for some of the variables. The assignment of a personal or unique identification number to each individual enumeration unit, and the use of these in administrative and

statistical registers, makes the matching significantly easier, faster and less costly.

6-8 No new administrative data source should be integrated into the census process without prior feasibility research. The quality of a data source may be established by acquiring 'test data' and assessing its quality at various stages in the census production cycles. This will aid the design of a census methodology that makes the most of the available administrative data and considers the impact of its use on the quality of the census overall.

6-9 Regardless of the data collection methodology, assessing the quality of the output of census data has always been an important and necessary task. There are several internationally accepted dimensions that can be applied to aid NSOs in their assessment of the quality of census data and processes to enable both producers and users to understand the quality of a statistical output. These key dimensions are (in no definitive order of priority):

- relevance,
- accuracy and reliability,
- timeliness and punctuality,
- coherence and comparability, and
- accessibility and interpretability.

6-10 There are several different ways and methods to assess the quality of statistics and particularly the quality of census outputs. It is always useful if more than one method is adopted. It is often the case, however, that assessing the quality of the census data produced using administrative sources presents a particular challenge due to the complexity and multi-dimensionality of the data used: indeed, many factors affecting quality are not quantitatively measurable. Moreover, being 'fit for purpose', means, by definition, that what high quality statistics should look like will necessarily vary from one user to another (for example while some users will regard accuracy as being of prime importance, others may prioritise timeliness and accessibility).

6.2 Recommendations

6-11 NSOs will need to plan and design any programme to transition from a traditional census to one that adopts a register-based (or combined) methodology in full accordance with national requirements and circumstances (for example, in utilising the available method and process for establishing unique identifiers for the purposes of linking records). There is no one size of

design that fits all. However, this Handbook concludes by recommending a number of key activities and processes that NSO would do well to attempt to follow in order to make any such transition as smooth and effective as possible, and to ensure high quality statistical output.

6-12 The NSO should in the earliest stages of planning, seek to **identify the administrative sources that may be relevant to their census**, set against a range of different user requirements. It is important to set out what the expected or required outcomes of using each source would be, against which an assessment of relevance can be made. This should include improvements to the efficiency of the census operation, particularly in terms of reductions in cost and respondent burden; improvements to the quality of the census; or the delivery of new or enhanced census outputs. Central to such assessment is setting out what the administrative source needs to deliver in terms of the target population and the required variables relating to this population.

6-13 The relationship between the NSO and the administrative data supplier should be supported by robust mechanisms of communication, written agreement and an excellent understanding of the needs of both parties. There must, by necessity, be a legal basis for the supply and use of the data. To help build the relationship and secure the data supply, the NSO should identify areas of benefit to the supplier, and to develop a feedback mechanisms to help the supplier better understand and, ultimately, improve, the administrative data.

6-14 **The NSO should engage with the supplier to gain an in-depth understanding of the data source.** Since administrative data are not collected for the needs of the census, it is important for the NSO to understand and assess differences between the required populations, concepts, definitions and time-related dimensions. This will usually require access to clear and comprehensive metadata about the administrative source. The metadata will provide a useful reference both for the census and for any other statistical programmes that might benefit from use of the source. More generally, a thorough assessment of the coherence and comparability of the administrative source, along with its limitations across the various quality dimensions, is essential. This includes the linkability of the source if this is a requirement for use in the census

6-15 It is important that the public and data users understand how and why administrative data are being used in the census. The NSO should, therefore, be transparent about the use by providing a clear justification of the benefits set against any risks and costs, and, in particular, by making unequivocal assurances

about maintaining confidentiality. This can be achieved through **good communication, including the publication of the procedures and policies in place that support the effective use and protection of data.**

6-16 The transition of integrating administrative data sources into the census process should be preceded by adequately resourced feasibility research that provides a 'proof of concept'. It is advisable that the NSO **carries out a number of test runs (using real data) well in advance of the main census to ensure any unforeseen issues are identified**, allowing enough time to correct or adjust the methods, processes or systems

6-17 **The NSO should record and publish the results of any quality assessment and assurance of all stages throughout the census production.** This will enable producers and users of the census to assess and provide appropriate feedback. This is important to order to ensure that users understand the strengths and limitations of the data, and can help determine whether the right balance has been achieved across the dimensions of quality in order to inform the improvement of future censuses

6-18 In conclusion, the over-riding recommendation is that any move from a traditional to a register-based census methodology **should represent only one element of the NSO's long-term vision for a developing a more general register-based statistical system.** To achieve such a vision it should be recognised that while NSOs generally have a long-standing expertise in information management, establishing a register-based statistical system will require statistical expertise, particularly in the linkage and matching of data in general. Accordingly, **it will be necessary to develop beforehand a full range of specific skills.** These include relationship management, information management, data analysis and associated subject matter skills. Specialized statistical skills may also be needed to support linking processes.

6-19 Moreover, as the UNECE has observed¹⁵, it should be emphasised that, though a register-based statistical system may clearly be the ultimate goal when considering the greater use of administrative data for statistical purposes - and particularly so for the census - in many countries it may seem a very distant goal, perhaps not attainable for many years. Nevertheless, by adopting a strategic plan based on step-by-step improvements towards creating the necessary pre-conditions, it will be possible to move gradually closer to this goal.

Annex A Definitions of terms and concepts

The following terms, concepts and definitions are adopted in this Handbook

Accessibility: The dimension of quality that is defined generally as the ease with which users are able to access the data.

Accuracy: The dimension of quality that refers to the degree to which the information correctly describes the phenomena it was designed to measure. More simply put, accuracy is the proximity between an estimate and the unknown true value.

Address: A number or similar designation that is assigned to a dwelling, business or any other structure. As well as serving postal delivery, addresses are also important for both administrative and statistical purposes, such as in civil registration systems and the census.

Address register: A register of residential addresses, often used for the purposes of creating enumeration areas comprising comparable numbers of dwellings. In cases of multi-occupied dwellings there can be more than one dwelling under a given residential address.

Administrative data: Data held on registers and other administrative sources relating to information collected by government and/or other organizations primarily for administrative (not research or statistical) purposes, such as registration, transaction and record keeping, usually for the provision of public services.

Administrative (data) source: A data holding that contains information collected primarily for administrative (not research or statistical) purposes. Such sources include administrative registers (with a unique identifier) and possibly other administrative data without a unique identifier.

Administrative register: A systematic collection of unit-level data organized in such a way that updating is possible (where 'updating' is the processing of identifiable information with the purpose of establishing, bringing up to date, correcting or extending the register. Such registers are primarily used in an administrative information system in which the data are used in the production

of goods and services in public or private institutions or companies. Administrative registers used for statistical purposes are normally operated by the state or jointly by local authorities, but some registers operated by private/commercial organizations may also be used.

Administrative units: The units for which administrative data are recorded. These may or may not be the same as those required for the statistical output (which are referred to as statistical units).

Attribute: A socio-demographic or economic characteristic relating to an administrative or statistical unit for which information is required for the purpose of the census.

Banding: The merging of variable categories into broader groups for the purposes of protecting the confidentiality of small cell data in outputs; a common example is the banding of single years of age into quinary age-groups.

Base registers: Those registers (sometimes referred to as ‘key registers’) that hold the basic information relating to the stock of the population or the set of other key units at any given time. Examples of base registers are population registers and dwelling or building registers. Such registers can be either: *administrative base registers* kept as a basic resource for public administration with the function of keeping stock of the population and to maintain identification information; or *statistical base registers* that are based on the corresponding administrative registers and whose principal functions are to define important populations and contain links to other base registers.

Benchmarking: Comparing data, metadata or processes against a recognised standard.

Big data: Large, often unstructured data sets that are available, potentially in real time, but which are difficult both to process efficiently and quality assure using traditional methods and technologies. The amount and variety of data available is growing rapidly, and such data sets are available in many formats, including audio, video, computer logs, purchase transactions, sensors and social networking sites. Some of these data are freely available on the web, whereas others are held by the private sector to which there may be no free access.

Building: Any independent structure containing one or more dwellings, rooms or other spaces, covered by a roof and enclosed within external walls or dividing walls which extend from the foundations to the roof, whether designed for residential or for agricultural, commercial, industrial or cultural purposes or for the provision of services. Thus, a building may be a detached house, apartment building, factory, shop, warehouse, garage, barn, etc.

Building (and/or dwelling) register: Registers usually held by land and property valuation agencies, land survey authorities or by local or municipal authorities responsible for the development of housing policies and urban planning that provide information on the location of all places where people might live as well as some of the core census housing topics.

Business (or Enterprise) register: A register of business organisations that can provide the location of places of work and information on the industrial sector in which the workforce is employed. A statistical business register has a different purpose than a company register. While a commercial/trade register serves a purpose of protection, accountability and control, a statistical register plays a central part in a system of official economic statistics.

Census day: The date of the reference period for the census as a whole, irrespective of when the data are collected.

Census estimates: A term used by some countries to describe the census output data to reflect the fact that the published figures do not purport to be true counts and that there must always be some degree of uncertainty (however small) in the accuracy of the numbers.

Clarity: The dimension of quality that relates to the availability of any supplementary information or metadata that may be necessary of help the user to interpret and understand the accompanying data.

Coherence: The dimension of quality that refers to the degree to which data that are derived from different sources or methods, but refer to the same topic, are similar.

Combined census: A census based on a combination of data taken from administrative registers and collected on questionnaires.

Comparability: The dimension of quality that refers to the degree to which data can be compared over time and domain.

Core topics: Those topics or data items that are regarded as being essential for providing key census information that meets most national and local user requirements.

Coverage: Coverage refers to the proportion of the population that a register or any statistical output aims to cover. *Over-coverage* is the error due to the inclusion in a sample or an administrative dataset of elements that do not belong there. *Under-coverage* is the failure to include required units in the frame, which results in the absence of information for those units.

Data controller: See 'Register holder'.

Data editing: The process by which data that exhibit errors, logical inconsistencies and spurious values are detected and corrected.

Derived variable: A new variable formed by using the data from other variables.

DSE (Dual System Estimation): A statistical method, based on a capture-recapture technique, applied to estimate the size of a population.

Dwelling: A room or suite of rooms - including its accessories, lobbies and corridors - in a permanent building or a structurally separated part of a building which, by the way it has been built, rebuilt or converted, is designed for habitation by one household all year round. A *dwelling* can be either a one-family dwelling in a stand-alone building or detached edifice, or an apartment in a block of flats. (See also 'Housing unit'.)

Dwelling register: See 'Building register'

Encryption: The process of converting information or data into a code for the purpose of preventing unauthorised access or disclosure.

Enterprise register: See 'Business register'.

Estimates: The term is sometimes used in this report to refer to the statistics produced in census outputs, and reflects the processes undertaken by NSIs to

adjust the input data to take account of under- or over-coverage, errors, missing counts and measures to control statistical disclosure.

Field enumeration: The process of collecting information on individual persons, households and/or housing unit covering the whole population (or a sample of it) using questionnaires

Frame: Any list, material or device that delimits, identifies, and allows access to the elements of the target population. A statistical register is a specific example.

Housing unit: A separate and independent place of abode intended for habitation by a single household, (sometimes referred to as a 'dwelling') or one not intended for habitation but occupied as living quarters by a household at the time of the census. Although intended for habitation by one household, a housing unit may, at the time of the census, be occupied by one or more households.

Imputation: The process by which missing input data items are replaced with plausible and consistent values.

Input data: The data (sometimes referred to as 'raw data') derived from an administrative source, before any processing or validation by the NSO.

Input quality: The quality of the raw administrative data as it is supplied to the NSO by the administrative authority

Interpretability: See 'Clarity'.

Key registers: See 'Base registers'.

Linkability: The ability to link data from several different administrative data sources to the same unit, usually by means of a unique identification number or code.

Macrodata: Aggregated data, generically referred to as 'statistics'.

Metadata: Data that describe or define other data. This broadly refer to anything that users need to know to make proper and correct use of the real data, in terms of accessing, processing, interpreting, analysing and presenting the information. Metadata include, for example, file descriptions, codebooks, processing details, sample designs and fieldwork reports. Metadata should be distinguished from 'Paradata' which generally refer to the details that describe the process by which the census data are collected, either from admin sources or a field enumeration/survey.

Microdata: Data relating to individual units. Within the context of the handbook the term is used to refer to both: (a) the information in a register relating to a single entry; and (b) anonymised data on the characteristics of units of a population, such as individual persons, households or establishments, collected in the census.

Non-core topics: Those topics or data items whose inclusion in the census may, in general, be considered less vital than the core topics, though, because of particular national circumstances and requirements, may regarded by the NSO as equally important.

Output data: The processed data as it is used in statistical outputs.

Output quality: The quality of the processed data as it is used in statistical outputs.

Over-coverage: See 'Coverage'.

Padrón (Spain): The Spanish population register, usually compiled for each Municipality.

Paradata: See 'Metadata'.

Periodicity: Within the context of the supply of administrative data, this is the time period between reference dates for consecutive input datasets. For the census more generally, it is the time between the dates of consecutive censuses (Census days).

Perturbation: The process of falsifying the data before publication by introducing an element of error for the purpose of preserving confidentiality. This error can be inserted either in the cell values after the table is created, when it is referred to as ‘output perturbation’, or the in the input data at the unit record level, referred to as data perturbation.

Statistical population register: A statistical register and a frame of persons usually resident (however defined) in a given country. Additionally, it often provides some demographic characteristics of individuals.

Privacy Impact Assessment (PIA): A process which assists organizations in identifying and managing the risks to privacy arising from new projects, initiatives, systems, processes, strategies, policies and business relationships.

Process quality: The effect of changes to the quality of data being used for the purpose of the census during the processing of the raw data by the NSO.

Punctuality: The dimension of quality that relates, when referring to data, to the time lag between the planned (and often pre-announced) publication dates and actual publication dates. In the context of the administrative source, it relates to the time lag between the expected (or contracted) date of the delivery of the data to the NSI and the actual date of delivery.

Raw data: See ‘Input data’.

Register: A systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, updating, correcting or extending the register.

Register-based census: A census where all data is collected from administrative registers. A census based on combination of data taken from registers and questionnaires is called a ‘combined census’.

Register keeper: See ‘Register holder’.

Register holder: The authority responsible for keeping and maintaining an administrative register (also referred to as the 'Register keeper' or 'Data controller').

Regularity: See 'Periodicity'

Relevance: The dimension of quality which, when referring to data, refers to the degree to which they meet the needs of users in terms of coverage and content. When referring specifically to data sources, the dimension refers to the degree to which such sources contain data that meets the needs of the NSO with respect to their intended use.

Reliability: The dimension of quality that refers to the degree of closeness of data values to earlier or subsequent data.

Rolling census: An alternative approach to the traditional model of census taking by means of a cumulative continuous survey, covering the whole country over a period of time, rather than on a particular day. There are two main parameters to consider in a rolling census: (a) the length of the periodicity, which itself is linked to the frequency of updating required; and (b) the sample size, which depends on the budget and the level of geographical analysis required for dissemination.

Secondary register: See 'Statistical register'.

Signs of life: An indicator used to minimise the over-coverage of persons recorded on different administrative registers derived by applying strict criteria or 'activity rules' to ensure that only living individuals who are usually resident are included in the census estimates.

Source quality: The quality of administrative sources from which data is supplied to NSO for the purpose of the census.

Specialised register: A register (sometimes referred to as a supplementary or subject-matter register) that contains information relating to the characteristics of unit records of interest that are not otherwise available in the base registers.

Statistical disclosure control: The process(es) by which the raw data taken from an administrative source or collected in the field is modified during data

processing in order to avoid the disclosure of information about identifiable individual persons or households.

Statistical register: A register processed for statistical purposes. A statistical register could be based on one or several administrative registers. Statistical registers are also referred to as 'secondary registers'.

Statistics: Numbers made from aggregation of micro or macrodata and presented in a table, graph or as a single number.

Subject-matter register: See 'Specialised register'.

Supplementary register: See 'Specialised register'.

Target population: The universe for which information is required. The target population is the set of the statistical units.

Timeliness: The dimension of data quality that refers to the lapse of time between the period to which the data refer (in the case of census data this is usually Census Day) and the date of publication of the data. In the use of administrative data, timeliness also refers to the length of time between the date of the event recorded in the data source and the date when the data are delivered to the NSO.

Top-coding: A process by which an upper limit is imposed to the published value of a variable (such as an age value of '90 or over'). Any value greater than this upper limit is replaced by the upper limit or is not published at all. Similarly, a bottom-coding is the process by which a lower limit is imposed on all published values for a variable.

Under-coverage: See Coverage.

Unit: The smallest entity to which any administrative data item refers. For the purpose of the census, units may refer to individual persons, households, buildings or dwellings.

Acronyms used in this Handbook

CAXI: Computer assisted multi-mode interview (Poland)

CES: Conference of European Statisticians

CRS: Civil registration system

CRVS: Civil registration and vital statistics

ESS: European Statistical System

EU: European Union

FPC: Precensal file (Spain)

GCC: Gulf Cooperation Council - a regional inter-governmental political and economic union consisting of Arab states of the Persian Gulf - Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates.

ILO: International Labour Organisation

INSEE: The French National Institute of Statistics and Economic Studies

LMS: Legal marital status

NSI: National Statistical Institute

NSO: National Statistical Office

OECD: Organisation for Economic Cooperation and Development

ONS: The United Kingdom's Office for National Statistics

PIN: Personal identification number

Annex B References

1 United Nations Department of Economic and Social Affairs, Statistics Division. *Principles and Recommendations on Population and Housing Censuses, Revision 3*. United Nations, New York, 2017.

https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf

2 United Nations Economic Commission for Europe. *Guidelines in the use of registers and administrative data for population and housing censuses*. United Nations, New York and Geneva, 2018.

<https://unece.org/DAM/stats/publications/2018/ECECESSTAT20184.pdf>

3 United Nations Economic Commission for Europe. *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing*. United Nations, New York and Geneva, 2015.

https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf

4 United Nations Economic Commission for Europe. *Measuring population and housing: practices of UNECE countries in the 2010 round of censuses*. United Nations, New York and Geneva, 2014.

https://unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

5 United Nations Economic Commission for Europe. *Register-Based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. United Nations, New York and Geneva, 2007.

https://unece.org/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

6 United Nations Department of Economic and Social Affairs, Statistics Division. *Handbook on the Management of Population and Housing Censuses, Revision 2*. United Nations, New York, 2017.

https://unstats.un.org/unsd/publication/seriesF/Series_F83Rev2en.pdf.

7 United Nations Department of Economic and Social Affairs, Statistics Division. *Guidelines on the Use of Electronic Data Collection Technologies in Population and Housing Censuses*. United Nations, New York, 2017

<https://unstats.un.org/unsd/demographic/standmeth/handbooks/data-collection-census-201901.pdf>.

8 United Nations Department of Economic and Social Affairs, Statistics Division. *Handbook on Measuring International Migration through Population Censuses*. United Nations, New York, 2017.

<https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/international-migration/2017-draft-E.pdf>

9 United Nations Department of Economic and Social Affairs, Statistics Division. *Handbook on Population and Housing Census Editing, Revision 2*. United Nations. New York, 2019.

<https://unstats.un.org/unsd/demographic-social/meetings/2019/newyork-egm-census/handbook-draft.pdf>

10 United Nations Department of Economic and Social Affairs, Statistics Division. *Report on the results of the UNSD survey on 2020 round population and housing censuses*. Report presented to the 51st Session of the United Nations Statistical Commission, 3-6 March 2020, New York.

<https://unstats.un.org/unsd/statcom/51st-session/documents/BG-Item3j-Survey-E.pdf>

11 United Nations Economic Commission for Europe. *Measuring population and housing: practices of UNECE countries in the 2000 round of censuses*. United Nations, New York and Geneva, 2008.

https://unece.org/DAM/stats/publications/Publication_on_2000_censuses.pdf

12 United Nations Economic Commission for Europe. *Measuring population and housing: practices of UNECE countries in the 2010 round of censuses*. United Nations, New York and Geneva, 2014.

<https://unece.org/measuring-population-and-housing>

13 United Nations Economic Commission for Europe. *Measuring population and housing in Eastern Europe, Caucasus and Central Asia: Review of practices in the 2010 round of censuses*. United Nations, New York and Geneva, 2016.

<https://unece.org/statistics/publications/measuring-population-and-housing-eastern-europe-caucasus-and-central-asia>

14 United Nations Economic Commission for Europe. *Keeping Count: Conducting the 2020 round of population and housing censuses during the COVID pandemic*. United Nations, Geneva, 2021.

<https://unece.org/statistics/publications/keeping-count>

15 United Nations Economic Commission for Europe. *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. United Nations, New York and Geneva, 2011.

https://unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf

16 United Nations Economic Commission for Europe. *Guidance in Modernizing Statistical Legislation*. United Nations, Geneva, 2018.

<https://unece.org/DAM/stats/publications/2018/ECECESSTAT20183.pdf>

17 United Nations Economic and Social Council. *Fundamental Principles of Official Statistics*. Resolution adopted by the Economic and Social Council on 24 July 2013 [on the recommendation of the Statistical Commission (E/2013/24)].

<https://unstats.un.org/unsd/dnss/gp/FP-Rev2013-E.pdf>

18 United Nations Economic and Social Commission for Asia and the Pacific. *Population Registers: A Key Resource for Producing Vital Statistics*. ESCAP Statistics Division, Stats Brief, No. 26 October 2020.

https://www.unescap.org/sites/default/d8files/knowledge-products/Stats_Brief_Issue26_Oct2020_A_Key_Resource_for_Producing_Statistics.pdf

19 Anders Wallgren and Britt Wallgren. *Register-based Statistics: Statistical Methods for Administrative Data*. 2nd edition, John Wiley & Sons Ltd, 2014, ISBN 978-1-119-94213-9.

20 International Labour Organisation. *International Standard Classification of Occupations (ISCO-08)*. ILO, 2008.

<https://www.ilo.org/public/english/bureau/stat/isco/isco08/>

21 United Nations Department of Economic and Social Affairs. *Methodology and Evaluation of Population Registers and Similar Systems*. Studies in Methods Series F No. 15. United Nations, New York, 1969.

https://unstats.un.org/unsd/publication/SeriesF/Seriesf_15e.pdf

22 United Nations Legal Identity Expert Group. *United Nations Strategy for Legal Identity for All: Concept note developed by the United Nations Legal Identity Expert Group*. United Nations Statistical Division, June 2020.

<https://unstats.un.org/legal-identity-agenda/documents/UN-Strategy-for-LIA.pdf>

23 United Nations Economic Commission for Europe. Guidelines on Statistical Business Registers. *United Nations, New York and Geneva, 2015*.

https://unece-modl.dotsoft.gr/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf

24 United Nations Department of Economic and Social Affairs, Statistics Division. United Nations Guidelines on Statistical Business Registers. *United Nations, New York, 2020*.

https://unstats.un.org/unsd/business-stat/SBR/Documents/UN_Guidelines_on_SBR.pdf

25 United Nations Department of Economic and Social Affairs, Statistics Division. *Handbook on Measuring International Migration through Population Censuses*. Series F No. 115. United Nations, New York, 2020.

<https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Handbooks/international-migration/2020-Handbook-Migration-and-Censuses-E.pdf>

26 United Nations Department of Economic and Social Affairs, Statistics Division. *Post Enumeration Surveys Operational Guidelines: Technical Report*. United Nations, New York, 2010.

https://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual_PESen.pdf

27 Raj Gautam Mitra, Martin Bratschi and Gloria Mathenge. *Population Registers: Definitions and Conceptual Framework*. Knowledge Series No. 1. Pacific Community, 2021,

https://sdd.spc.int/digital_library/knowledge-series-1-population-registers-definitions-and-conceptual-framework

28 United Nations Economic Commission for Europe. Guidelines for assessing the quality of administrative sources for use in censuses. United Nations, New York and Geneva, 2021.

https://unece.org/sites/default/files/2021-10/ECECESSTAT20214_WEB.pdf

29 Organisation for Economic Co-operation and Development, *International Migration Statistics for OECD Countries: Sources and Comparability of Migration Statistics*. OECD, 2001.

<https://stats.oecd.org/glossary/detail.asp?ID=2089>

30 UN Expert Group on Migration Statistics, Task Force 2: Key concepts and definitions related to international migration. *Final Report on Conceptual frameworks and Concepts and Definitions on International Migration*. 27 April 2021.

<https://unstats.un.org/unsd/demographic-social/migration-expert-group/task-forces/TF2-ConceptualFramework-Final.pdf>

31 United Nations Department of Economic and Social Affairs, Statistics Division. *Principles and Recommendations for a Vital Statistics System, Revision 3*, New York, 2014

https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/CRVS/M19Rev3-E.pdf

32 United Nations Department of Economic and Social Affairs, Statistics Division/United Nations Development Programme/United Nations Children Fund. *Implementation of the United Nations Legal Identity Agenda: United Nations Country Team Operational Guidelines*.

https://unstats.un.org/legal-identity-agenda/documents/UNCT_Guidelines.pdf

33 United Nations Statistics Division. *Technical Report: Measuring Sustainable Development Goals Indicators Through Population and Housing Censuses and Civil Registration and Vital Statistics Data*. (Version of 12 October 2020)

https://unstats.un.org/unsd/demographic-social/census/documents/tr_on_sdg_in_phc_crvs.pdf

34 Albert Kraler, Christina Hollomey and Alfred Wöger. *National Data Collection Systems and Practices: Country Report Austria*. PROMINSTAT, May 2009.

<http://www.prominstat.eu/drupal/?q=node/139>

35 World Bank Group. *ID4D Practitioner's guide: Version 1.0, October 2019*. World Bank, Washington, DC, 2019.

<https://documents1.worldbank.org/curated/en/248371559325561562/pdf/ID4D-Practitioner-s-Guide.pdf>

36 European Statistical System. The second revision of the quality assurance framework of the European Statistical System, 2019.

<https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>