Quality Guidelines

for Administrative based Population and Housing Censuses

in GCC Countries in the 2020 Round

2018

This quality reference was prepared according to GCC-Stat standards

GCC-Stat, 2018. Quality Guidelines for Administrative based Population and Housing Censuses in GCC Countries in the 2020 Round. 2018. *Muscat* - Oman.

All correspondence should be directed to:

GCC-Stat

P. O. Box 840, Postcode 133, Muscat – Oman.

Tel: +968 24346499

Fax: +968 24343228

E-Mail: info@gccstat.org

Website: http://www.gccstat.org

Preface

The Harmonised 2020 GCC Population and Housing Census is one of the key strategic statistical projects across the Gulf Cooperation Council (GCC). This reflects that the census is one of the most important statistical projects for any statistical office. Many government and private sector decisions make extensive use of statistics from the census. This means that the census data must be accurate, free from errors and inconsistencies.

However, quality census information must also be timely, relevant, accessible, coherent and consistent. Achieving these quality dimensions requires statistics offices to build quality control and quality assurance processes into all the phases of the census cycle.

All Statistical systems require an integrated approach to managing quality. Accordingly, GCC-Stat has prepared a Data Quality Framework for GCC Statistics.[1] Projects such as the 2020 Census require quality to be built into all stages. Accordingly, GCC-Stat has prepared specific Quality Guidelines for the GCC 2020 Administrative based Population and Housing Censuses.

The Census Quality guidelines are underpinned by the Data Quality Framework for GCC Statistics and international best practices as set out in relevant United Nations Census manuals and guidelines. Manuals and guidelines related to the use of Administrative data also provided valuable references. Experiences from Nordic countries, United Kingdom, Canada, Netherlands and New Zealand have also informed the guidelines.

While the Census Quality Guidelines focus on Administrative based censuses, these guidelines also apply to Fieldwork (Traditional) and Combined Censuses, which use both Administrative and Field work methodologies.

Feedback from participants in a GCC regional census workshop in 2017 was used in finalising the guidelines. GCC-Stat thanks participants at that workshop as well as colleagues in GCC-Stat who have prepared the publication.

---

[1] إطار ضمان جودة البيانات الإحصائية لدول مجلس التعاون الخليج لدول العربية (Data Quality Framework for GCC Statistics)
https://www.gccstat.org/images/gccstat/docman/Standards/b00k.pdf.GCC-Stat, 2018.

Contents

Introduction

To be useful, census information must be accurate, timely, relevant, accessible, coherent and consistent. Achieving these quality dimensions requires statistics offices to build quality control and quality assurance processes into all the phases of the census cycle. Countries in the GCC are also moving to make more use of administrative data, including for the 2020 Census round. This creates a new set of requirements in managing quality.

GCC-Stat has therefore produced these guidelines for managing quality in administrative censuses. They are part of a series of guidelines to assist countries in the 2020 Census round, and are underpinned by the GCC Quality Framework prepared by GCC-Stat and published in 2018.

These Census guidelines focus on the implementation of Quality Management through all the phases of the census cycle, with particular emphasis on the special requirements for administrative based censuses.

In using administrative data, NSOs need to:
- understand the data;
- identify any errors, uncertainty or bias in the data;
- make efforts to understand why the errors occur and to manage them,
- determine whether the administrative data is suitable for statistical purposes and
- communicate to users how the use of administrative data could affect the statistics and their use.

The guidelines therefore include detailed information to help countries assess the quality of administrative sources. This includes a detailed checklist to be used in assessing the quality of administrative records in the census.

Section 1 describes the General Quality framework that underpins quality statistics in the GCC, in the context of the international quality frameworks. Section 2 describes how this framework is applied all the stages of the 2020 Census.

Administrative censuses have additional quality considerations. Some of the specific issues are discussed in Section 3. This section also introduces the GCC Administrative Census Checklists that can be used in the assessment and evaluation of administrative sources.

Finally, Section 4 sets out the framework for Quality Reporting, including the Census template for reporting.

There are also three Appendices. Appendix 1 sets out the Administrative Census assessment tool. Appendix 2 provides a template for quality reporting on the Census. The guidelines conclude with Appendix 3, which provides a list of additional resources.

Abbreviations
GSBPM – Generalised Statistical Business Processing Model
NSO – National Statistical Office

1 General Quality Framework

Countries across the GCC are increasingly concerned with the management of quality in their statistical processes. GCC- Stat has published the GCC Quality framework - إطار ضمان جودة البيانات الإحصائية لدول مجلس التعاون لدول الخليج العربية. This framework, which is based on the standard international frameworks established by the UN, IMF and Eurostat, addresses both quality control and quality assurance across the quality dimensions of relevance, accuracy, timeliness and punctuality, accessibility and clarity; and comparability and coherence.

The GCC Quality framework and the associated quality dimensions are used throughout these guidelines.

1.1 International Quality Frameworks

In the context of a national statistical office, Quality is defined in terms of **Quality Components (or Dimensions).** The Quality Management System is typically expressed in the form of a **Quality Assurance Framework.**

Some NSOs have adopted ISO 9001 Quality Management Systems as their overall quality framework. Others, like Statistics Canada, have constructed a quality management system, tailored to their particular needs. A number of international and regional organisations, such as the IMF Statistical Division, the UN and Eurostat have also prepared Statistical quality frameworks. The most common international standards are

- UN Fundamental Principles of National Official Statistics
- UN National Quality Assurance Framework (NQAF)
- IMF Data Quality Assessment Framework (DQAF)
- European Statistical System (ESS)

While the GCC Quality Framework is mainly based on the UN National Quality Framework (NQAF), it also draws on the other international standards[2].

Output Quality and Quality of Processes

While these international quality frameworks have a strong focus on quality assurance and output quality, they also recognises that ***Output quality*** is achieved through ***process quality.*** Process quality has two broad aspects

- Effectiveness: which leads to outputs of good quality
- Efficiency: which leads to production of outputs at minimum cost to NSO and to respondents

A common aspect of these is the need to manage both the quality of the processes used to prepare statistics, as well as manage the quality of the outputs. All of this also needs to be conducted within the context of the overall framework for official statistics – the UN Fundamental Principles of National Official Statistics.

The procedures set out in these guidelines therefore considers both process and output quality.

---

[2] More information on the international standards is available in the GCC Quality Framework.

Quality Control and Quality Assurance

Both Quality Control and Quality Assurance have key roles in the management of statistical quality.

Quality Control

Quality Control provides regular and consistent checks to ensure data integrity, correctness and completeness; and also identify and address errors. Examples of Quality Control are the physical checking of completed questionnaires and data editing.

The aim of **Quality Control** is to deal immediately with any substandard data, either by fixing or by discarding it. In addition, **Quality Control** checks are conducted to ensure the accuracy and completeness of the program plans, including all schedules and cost estimates, agreements (e.g. memoranda of understanding) and contracts. Quality control checks are also conducted on the design (the mix of administrative and fieldwork data.) Testing (discussed in Section 2) is also a key form of Quality Control.

Quality Assurance

These activities test the accuracy and reliability of the processes. They aim to provide confidence in the quality of the product by assessing the performance of a process according to certain criteria. Examples include quality audits, and reviewing of performance measures and quality indicators after the survey.

While Quality Control has a very short timeframe focus, **Quality Assurance** has a longer time horizon. It aims to fix identified problems in the next cycle of the activity – for example the next collection.

In the case of the census, Quality Assurance also provides feedback on the previous phase of the census cycle, including identifying areas for future improvement.

Using Quality Control and Quality Assurance

Achieving a successful census therefore requires both Quality Assurance and Quality Control. Quality Assurance is focused on Prevention, including helping to ensure that these errors are not repeated in future censuses. Quality Control is focused on Detecting Errors and Correcting them. They need to work together and so the Census guidelines include both Quality Control and Quality Assurance elements.

1.2 GCC Statistics Quality Assurance Framework

The GCC Data Quality Assurance Framework (GCC DQAF) has 19 elements as shown in Figure 1.

Figure 1: GCC Data Quality Framework

Managing the Statistical System	Managing the Institutional environment	Managing Statistical Processes	Managing Statistical Outputs
1. Coordinating the National Statistical System	4. Assuring professional independence	10. Assuring methodological soundness	14. Assuring relevance
2. Managing relationships with data users and data providers	5. Assuring impartiality and objectivity	11. Assuring cost-effectiveness	15. Assuring accuracy and reliability
3. Managing statistical standards	6. Assuring transparency	12. Assuring soundness of implementation	16. Assuring timeliness and punctuality
	7. Assuring statistical confidentiality and security	13 Managing respondent burden	17. Assuring accessibility and clarity
	8. Assuring the Quality Component		18. Assuring coherence and comparability
	9. Assuring adequacy of resources		19. Managing metadata

The GCC framework is designed to be a tool for GCC countries to prepare their own national (or in some cases sub-national) quality assurance plans. It is expected that the framework will improve and expand as new practices are developed by GCC countries. Therefore, the framework is concerned not just with specific projects or statistical outputs such as the Census, but the broader management of the statistical system.

Main Dimensions of Quality in GCC Statistical Outputs

The dimensions of quality that apply to all GCC statistical outputs are:
- Relevance
- Accuracy
- Timeliness and Punctuality
- Accessibility and Clarity
- Coherence and Comparability

Relevance means that the final statistics, including timing and format are **relevant** to data users.

Accuracy means that the final statistics do not contain errors. While some errors may remain, these should not affect the main uses of the statistics.

To be useful, statistics must be published in a timeframe that allows the information to be used. **Timely** statistics should be published as close as possible to the reference date. A

related aspect is **Punctuality** – delivering statistics to users according to the agreed delivery dates.

Accessibility and **Clarity** are dimensions related to the ability of users to access necessary information in formats that meet their needs. This includes availability of information, suitability of the dissemination format, availability of metadata, and whether the user has information to know what statistics are available and how to access that information.

Coherence reflects whether the data can be combined with other statistical information within an integrated framework over time. The use of standard concepts, definitions and classifications promotes coherence. Equally important is internal coherence of data across the statistics.

Statistics are most useful when they enable reliable comparisons, such as between countries or between regions within a country, and over time. Standard concepts, definitions and classifications are important enablers of **Comparability**.

2 Quality in Population and Housing Censuses

2.1 Overview

Population and Housing Censuses in the GCC include all the steps or phases of the census:

- planning and monitoring
- collection
- processing and analysis,
- evaluation,
- dissemination

> The quality of the census needs to be managed across all of these phases. Focusing only on one area, e.g. collection means that errors introduced in other phases will not be identified and corrected. This means that the resulting statistics will be of poor quality with consequences for decisions.

Census Cycle

All censuses follow a standard cycle, with each phase dependent on the previous phase, as shown in Figure 2. The quality of the output from any phase has a direct effect on the success of the next phase and ultimately the overall project. Planning and Monitoring have a critical role in linking the cycle together.

Planning and Monitoring[3]

These are the project management activities such as project planning, budgeting, project monitoring and controlling. These support the technical and operational work, to help ensure that the project meets its objectives.

Preparation

This phase covers all the preparation activities. (In this model, the Preparation phase includes the Generalised Statistical Business Processing Model (GSBPM) [4] phases of Specify Need, Design and Build.) These activities cover identification of requirements, as well as the detailed technical and statistical work to define and prepare the statistical outputs, concepts, methodologies, collection instruments and operational processes. This phase also includes the activities necessary to build and test the statistical processes and systems, including the supporting IT systems and infrastructure.

At the conclusion of the preparation phase, the project is ready to go live.

Field Collection/Data Collation

This phase, also known as Enumeration, covers the activities involved in the collection of data. This may occur directly through fieldwork or indirectly using data already recorded in administrative registers.

---

[3] See also *Guideline to Planning and Preparing Censuses in GCC Countries*, GCC-Stat, 2018

[4] For more on GSBPM, see *Generic Statistical Business Process Model, GSBPM, Version 5*, December 2013, United Nations Commission for Europe, http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5

Processing and Analysis

This phase involves the cleaning of data, preparation for analysis and then the subsequent analysis. **Processing** includes a number of sub-processes that check, clean, and transform input data, including coding and editing. **Analysis** involves the production of statistical outputs, detailed examination and preparing the statistical reports for dissemination.

The Processing and Analysis activities are linked and dependent on one another. Investigations undertaken in Analysis may identify the need for additional processing. Additional analysis may then be required.

Dissemination

This phase covers all the activities associated with the release of the statistical products to customers and clients. These activities include support for customers to access and use the outputs.

Evaluation

The focus here is on the activities needed to evaluate the overall quality of the results from a statistical perspective. The activities may include, for example, preparing, conducting and analysing a coverage assessment study such as a Post Enumeration Survey, and/or conducting Demographic analysis or Matching studies. (There is also a clear link to the project management evaluation and closure activities included in **Planning and Monitoring**.)

Figure 2: Census Cycle
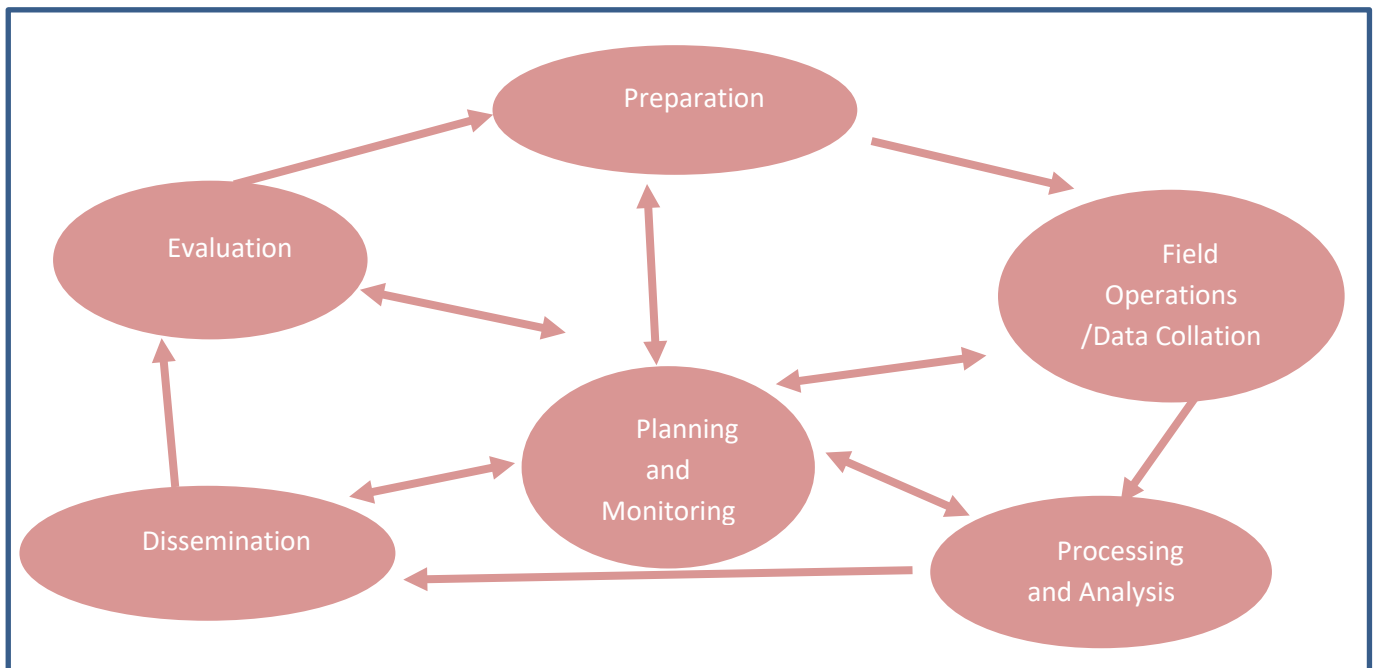
Census Collection Methodologies

As Figure 3 shows, there are a number of possible methodologies for conducting Population and Housing censuses. These include:

1. Fieldwork (Traditional) Census - All persons and housing units provide data, either by self-completed form or to an enumerator.

2. Administrative (Register) Census – all data collected from existing administrative sources – no fieldwork involved
3. Combined Census using Administrative and Fieldwork methodologies
4. Rolling Census – where part of the country is enumerated every year, using a systematic sample
5. Administrative Census supplemented by sample surveys

Figure 3: Census Methodologies

Methodology	Key feature of Collection	Note
Fieldwork (Traditional census)	Obtained directly from respondents	May include enumerators collecting information by face to face or telephone interview, or self-completed census questionnaires (on paper or by Internet). May also include mail out/mail in, or a combination.
Administrative (Register) Census	Existing administrative sources are linked together at individual level	Sources include Registers for individuals, households and dwellings; other administrative registers such as business, tax, education, employment registers.
Combined Census using Administrative and Fieldwork methodologies	Data from administrative registers combined with data from one or more new surveys or full census field enumeration.	Register data may be used to prefill the questionnaires, with information verified or corrected during data collection. Another method is to use fieldwork to provide information on topics not available from registers, or to adjust data that are of poor quality in registers.
Rolling census	Collected through a continuous cumulative survey covering the whole country over a period of time (generally years).	Modelling generates estimates of detailed characteristics for different geographic levels, and time periods. The sample can be accumulated over time to produce statistics at the lowest levels of geographic detail.
Administrative Census supplemented by existing sample surveys	Administrative registers are combined with information from existing surveys. No additional field data collection takes place	Information from existing administrative sources is linked at the individual level with information from existing sample surveys (e.g. labour force survey, living standards survey).

Essential features of population and housing censuses

Irrespective of the census methodology, all censuses have a number of essential features: individual enumeration, universality within a defined territory, simultaneity, defined periodicity

and small area statistics[5]. Figure 4 describes the requirements and implications for each of these features.

Figure 4: Essential features of Population and Housing Censuses

Feature	Requirement	Implication
Individual enumeration	Each individual and each set of living quarters is enumerated separately and the characteristics are separately recorded	Irrespective of the collection method, each individual and address/housing unit must be uniquely identified
Universality within a defined territory	Should cover a precisely defined territory	Each person usually resident should be included. Every set of living quarters should be included.
Simultaneity	Each person and each set of living quarters should be enumerated for the same reference date. Collected data should refer to a well-defined reference period	The reference period will be the day of the census for most data items. However, for some data items, (e.g. labour force topics), it may be a period prior to the census. All population units should use the same reference periods for each topic.
Defined periodicity	Censuses should be taken at regular intervals so that comparable information is made available in a fixed sequence.	Countries in the GCC should conduct a census every ten years.
Universality within a defined territory	Should cover a precisely defined territory	Each person present and/or residing within this territory should be included. Housing census should include every set of living quarters irrespective of type.

Role of Managers in achieving quality

Managers have a vital role in achieving quality in the Census[6]. This includes:
- Establishing Quality Culture
- Ensuring user expectations are known and met
- Ensuring processes are documented and understood
- Appropriate problem solving techniques
- Effective Interdisciplinary project team

Establishing Quality Culture

The biggest challenge for managers is to establish a culture with a focus on quality issues and to obtain the commitment of staff to strive to achieve high-quality goals. International experience is that managers who do not delegate responsibility will find it difficult, if not impossible, to establish teams that strive for high-quality outcomes.

Managers have specific responsibilities in establishing a quality culture including:

---

[5] See *Principles and Recommendations for Population and Housing Censuses – Revision 3,* 2017
https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf
[6] Based on *Principles and Recommendations for Population and Housing Censuses – Revision 3* Section XIV – Quality Assurance.

a) using quality project management tools and techniques;
b) managing the project using good project management techniques[7]
c) managing project stakeholders;
d) creating the conditions for good team work;
e) ensuring staff have clear roles and responsibilities;
f) ensuring that the team has the right mix of skills and expertise;
g) delivering the project deliverables and benefits;
h) leading the project team;
i) evaluating and closing the project.

Meeting User Expectations

Managers need to ensure that users are properly identified and that their requirements and expectations are built into both planning objectives (e.g. deadlines) and into the required systems. Managers also need to establish feedback mechanisms on proposed topics, output products and services.

Documentation and Knowledge Sharing

All processes and systems, including systems and processes for Quality Assurance and Quality Control, need to be documented and clearly understood. Processes for managing quality need to address questions such as how quality will be measured, who is involved in identifying root causes of quality problems, and how process improvements are going to be implemented.

Appropriate Problem Solving Approaches

All census projects encounter problems. How Census managers approach the solving of these problems is recognised as the greatest test of management commitment to genuine quality improvement. An environment where the emphasis is on finding faults (rather than on finding solutions to problems), or on excessive competition, will assure that staff cease to be part of the solution and become part of the problem.

Managers need to take upon themselves the responsibility for problems, as they are ultimately responsible for the systems or processes that caused the problems. They should not seek to transfer the problems to lower-level staff.

However, there will be cases where individuals are justifiably responsible for negative impacts on quality. These individuals need to be dealt with decisively and consistently. Training and guidance should be provided, and then if necessary disciplinary measures administered.

Effective Interdisciplinary project team

Establishing an interdisciplinary project team will help ensure that quality considerations relating to all the census steps receive appropriate attention. All Census project teams need to be adequately staffed with the full range of expertise. Each of the following areas contributes to various quality dimensions.

Subject matter and classification experts bring knowledge of content, client needs, relevance and coherence and comparability. While **Methodologists** bring expertise in

---

[7] These include establishing a project management office, implementing risk management and change management techniques

statistical methods and data quality trade-offs, especially with respect to accuracy, timeliness and cost.

Experienced **Operational staff** bring expertise in operational methods, and a concern for practicality, especially related to efficiency, qualified field staff, satisfied respondents and implementation of operational quality control. **Systems experts (**i.e. **Information Technology and Management experts)** bring a systems view, and knowledge of technology standards and tools. They are able to automate processes, including quality control that help achieve targeted census timeliness and accuracy. **Dissemination experts** in collaboration with subject-matter experts will focus on census output accessibility and clarity.
Figure 5 shows the contributions of each of these areas.

Figure 5: Contributions of Expertise to the Census Cycle

Census Phase	Subject Matter and Classification Experts	Method-ologists	Operations Experts	Systems Experts	Dissemination Experts
Preparation					
Content Consultation	Major contribution	Contribution	Do not contribute	Contribution	Contribution
Design	Major contribution	Major contribution	Major contribution	Major contribution	Contribution
Form testing	Major contribution	Major contribution	Major contribution	Major contribution	Contribution
Collection					
Field Collection	Contribution	Do not contribute	Major contribution	Contribution	Contribution
Administrative Data Collection	Major contribution	Major contribution	Contribution	Major contribution	Contribution
Processing	Major contribution	Major contribution	Contribution	Major contribution	Contribution
Analysis	Major contribution	Major contribution	Contribution	Contribution	Contribution
Dissemination	Major contribution	Contribution	Do not contribute	Contribution	Major contribution
Evaluation	Major contribution	Major contribution	Contribution	Contribution	Contribution
Project Management	Major contribution	Major contribution	Major contribution	Major contribution	Major contribution

Key

■ (dark green)	Major contribution
■ (light green)	Contribution
□ (white)	Do not contribute

2.2 Census Quality Considerations

Reliability and Reputation of the NSO

The census is one of the most important statistical projects for any statistics office. The reliability and reputation of the NSO is often linked to the success of the census.

Decisions based on census data require users to have good quality information. Examples of poor quality information from a user perspective include:

- **Outputs** not meeting user requirements, e.g. missing specific combinations of data, or required level of geography
- **Output timetables not met** – For example, not available at a time that meets user requirements.
- **Inconsistent data** – This can include information from the Census not consistent with other statistics – as well as not specifying reasons for inconsistencies
- **Outputs not clear** – Examples include limited supporting metadata, unclear combinations of statistics, publications where the text or figures are unclear.
- **Outputs not accessible** – Examples include whether the access format is not appropriate to users (such as statistics only available in pdf format), as well as restrictions on access.
- **Outputs not comparable** - Census statistics not able to be compared to other statistics
- **Errors in published statistics** - Errors in the final numbers or text.

If the users have poor quality census information, this is likely to result in lack of trust for the Census or the NSO, and so harm the reputation of the NSO. Other factors that can harm the reputation of the NSO include:

- **High Levels of Non-response** – where overall response rate is lower than expected, or there are low response rates for individual questions; then users and decision makers will not have confidence in the NSO to provide accurate data
- **No measures of quality** – Where no quality measures are provided, some users will attempt to prepare measures themselves. Users may make wrong assumptions that could affect the NSO's reputation.
- **Lots of Re-work and Budget blowout** – If the census requires unexpected rework, then there are likely to be extra costs and/or delayed deadlines.

Applying Quality Dimensions throughout the phases of the Census cycle

The Quality Dimensions of Relevance, Accuracy, Timeliness and Punctuality, Accessibility and Clarity, and Coherence do not apply equally to all census phases, as Figure 6 shows. For example, Accuracy applies to all the phases of the Census cycle; however, Accessibility and Clarity apply mainly to the Analysis and Dissemination phases.

In addition, the focus differs across the census cycle. For example, Accuracy is important through all the phases, however Relevance has a Primary role during Preparation and then again during Dissemination. Timeliness and Punctuality are important considerations during the Preparation, Collection and Dissemination phases, but still have a role in the other phases.

Figure 6: Applying Quality Dimensions across the Census cycle

Quality Dimension	Census Phase					
	Preparation	**Collection**	**Processing**	**Analysis**	**Dissemination**	**Evaluation**
Relevance	Primary	Secondary	Secondary	Primary	Primary	Secondary
Accuracy	Primary	Primary	Primary	Primary	Primary	Primary
Timeliness and Punctuality	Primary	Primary	Primary	Primary	Primary	Secondary
Accessibility and Clarity	Secondary	Secondary	Secondary	Primary	Primary	Secondary
Coherence	Primary	Primary	Secondary	Primary	Primary	Primary

Note: In all cases, this is underpinned by **Planning/Monitoring** (i.e. Project Management).

Key

Primary Quality Role	
Secondary Quality Role	

Errors in the Census

All censuses have errors. Figure 7 shows examples of common errors, and the likely impact.

All of these errors are interlinked. A poorly implemented census will have many **Accuracy** errors. These include errors in the various operational activities (**Operational errors**). This will mean that the census will fail to count everyone correctly – **coverage errors.** Censuses with high level of coverage error will have higher non-response/higher levels of missing data (**content error)**, and are therefore likely to have **Coherence and comparability errors**. A poorly implemented census may not collect a full range of metadata to help explain the errors and the impact on the statistics. This will result in **Accessibility and Clarity errors.**

Extra work will be needed to attempt to reduce the obvious mistakes in the outputs. This in turn may result in the Census outputs not being published according to the release calendar (**Timeliness errors**), with the consequence that the final release is too late to meet needs of users. This then affects the **Relevance** of the census to users and the wider stakeholders.

Figure 7: Examples of Common Errors in the Census

Quality Dimension	Examples of Common Errors	Likely Impact
Relevance	Topics not Relevant Outputs not Relevant Questionnaire/ Classifications not relevant to Users	Outputs not Relevant to User
Accuracy	*Operational Errors.* Errors in Data collected by Field Staff Errors in Processing (coding, editing) Errors in Dissemination processes *Output Errors.* Errors in Output data Coverage Errors (missing/duplicate people/houses) Content Errors (Non-response, missing data)	Errors in Published Statistics
Timeliness and Punctuality	Census outputs not published according to release calendar	Output timetables not met
	Timing of release of data too late for users	Outputs not Relevant to User
Accessibility and Clarity	Explanations (metadata) not clear or not provided	Outputs not Clear to Users
	Dissemination methods not appropriate	Outputs not Accessible
Coherence	Census statistics not internally consistent	Inconsistent Statistics
	Census statistics inconsistent with other statistics	
	International classifications not applied	Outputs cannot be compared
	Budget exceeded	NSO Project Management questioned
	Unscheduled extra work	
	Different views of quality of census	Quality not Understood

While it is important to respond and correct the major errors, it is also critical to examine the underlying causes. Resolving these as early as possible can reduce further errors, and so improve quality. Figure 8 shows possible underlying reasons for common errors in censuses.

Figure 8: Reasons for Errors in Censuses[8].

Error Types	Common Errors	Possible Reasons
Operational Errors	Errors in Collected Data	Systems and Processes not well tested Poor Quality Control checks Poorly prepared questionnaires/ instructions Poorly managed field staff/ Wrong staff hired
	Errors in Processing	Poor Systems and Processes Poor Quality Control checks
	Issues arise after Testing	Testing not conducted properly
Coverage Errors	Missing/duplicate people/housing units	Enumeration poorly prepared Public not properly engaged
Content Errors	Non-response, missing data	Poor instructions Systematic errors not corrected in time
	Inconsistent and inaccurate data	Poor instructions/ Poorly tested questionnaires Poorly prepared/implemented editing and imputation systems Poorly trained or managed field/processing staff
Accessibility and Clarity Errors	Explanations (metadata) not clear or not provided	Metadata not collected/collated Metadata not clearly displayed Metadata standards not followed
	Dissemination methods not appropriate for users	Users not consulted on requirements
Timing and Punctuality Errors	Census outputs not published according to release calendar	Delays in Dissemination phase Dissemination phase not properly planned Dissemination Systems and Processes not properly implemented Release timetable not properly planned
	Data released too late for users	Users not consulted on requirements
Coherence Errors	Statistics not internally consistent	Census data not properly analysed Not enough time allowed for Analysis Analysis techniques not properly applied Comparability needs not considered in preparation phase
	Census statistics inconsistent with other statistics	
Relevance	Published Statistics not used	Statistics don't meet user needs Users poorly supported in use of statistics
	Questionnaire not relevant to Users	Poor mapping between outputs, topics and questionnaires

---

[8] Specific errors relating to Administrative Censuses are discussed in Section 3.

There are a number of common underlying themes, including:
- Quality not built in at the start/ Not implemented in all phases
- Poor planning and/or monitoring
- Limited or poor consultation with users
- Not enough time spent in preparation
- Processes and systems not properly prepared, tested or implemented
- Poor training
- Processes and systems not properly implemented
- Lack of feedback processes – between all levels and teams
- Not working as a team

These underlying reasons point to the need for Quality Management plans throughout all the census phases. Figure 9 is a checklist of activities that need to be included in these plans.

Census Phase	Activity	Quality Management Actions
Planning and Monitoring	Planning and Monitoring	Project Planning and Monitoring Processes
		Independent review of Project plans
		Change control
		Project Evaluation
Preparation	Training	Appropriate training for Technical and Statistical staff
		Standard training techniques, training material and manuals
		Feedback on all training
	User Consultation	Consultation with Users on Topics, Output requirements
	Design	Prioritisation process for Basket Topics and Outputs
		Independent review of design of Census
		Sign-off (approval) processes for key design decisions
		Independent checks that design is being implemented
	Testing	Detailed testing programme for all processes, procedures and systems
	Outsourcing	Briefings for all Out-sourced suppliers
		Feedback from Suppliers
Collection	Training	Training for all staff
	Operations	Clear roles
		Quality Control checks
		Staff Feedback Processes
Processing	Training	Training for all staff
	Operations	Clear roles
		Quality Control checks
		Staff Feedback Processes
Analysis	Training	Training for all staff
	Operations	Clear roles
		Quality Control checks
		Staff Feedback Processes
Dissemination	Training	Training for all staff
	Operations	Clear roles
		Quality Control checks
		Staff Feedback Processes
		User Feedback Processes
Evaluation	Preparation	Evaluation planned, prepared and tested
	Implementation	Statistical Evaluation conducted
		Evaluation results published

2.3 Testing and Trials

Role of Testing and Trials

In any census project, testing and trialing of processes, systems and methodologies is a key part of preparation.

Testing has a number of functions, including:
- Helping to determine if the proposed design will work as intended
- Key part of creating new processes and methodologies
- Identifying aspects which don't work as planned

In this way, testing helps to reduce risk, inform key decisions, operational planning and preparation, as well as providing training for the main census. Testing can also help to give an early assessment of whether the final statistics will be of the required quality.

Census testing should include all the phases of the census, including
- Preparation –such as questionnaire testing, as well as all IT systems
- Collection – for example testing collection methods and processes
- Processing – for example testing statistical methodologies, systems and processes, including links with Collection
- Analysis – such as testing of the analysis techniques and links with Processing
- Dissemination – for example testing of systems and processes, as well as statistical disclosure control.

Testing often requires all the skills discussed in Figure 5. The different team members bring different perspectives to this key stage of the census process.

Additional Testing in Administrative Census projects

When moving to an Administrative Register Census, testing is critical. Testing will help determine if the proposed methods for using administrative registers will work as intended.

However, Administrative censuses also mean new sources, processes and methods. This means that testing is also critical to help the NSO and users understand the impact of changing census methodologies. As the UK Office for National Statistics noted in 2016, "Given importance of accurate statistics, (it is) high risk to move straight to an administrative based census without benchmarking new methods [9]".

> **In the cases of GCC countries, this means that extensive testing is critical for countries who are moving to some form of Administrative or Combined Census.**

Testing of Register Sources and the Census Design

It is important to test whether the planned approach will work and whether the administrative records meet statistical requirements. This will take time and needs to be conducted carefully.

---

[9] See Office for National Statistics *Evaluating the potential for moving away from a traditional census*, paper for Conference of European Statisticians Group of Experts on Population and Housing Censuses Eighteenth Meeting Geneva, 28 - 30 September 2016,
https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2016/mtg1/CES_GE.41_2016_18E.pdf

The Nordic countries (Denmark, Sweden, Finland, Norway, and Iceland) were the first countries to move to administrative register censuses. They completed this stage in a systematic manner.

First, subject matter statistics were prepared from administrative registers. These were tested and published in different areas. This step gave the NSO a very good understanding of all the dimensions of quality and enabled them to get feedback from users and stakeholders. As soon as the NSO considered that the quality was sufficient to use in the census, the variables on the backbone registers (Population, Address, Establishment, etc.) were progressively introduced.

The move to totally register-based census could only proceed, if administrative register statistics had been developed for all the topics relevant for the census.

Based on this experience, the Nordic countries consider that the main factors that determine the timeframe to implement a fully register census are:
- information technology and
- suitability of the register records.

While developments in information technology have reduced the time between establishing administrative registers and their use in official statistics, the Nordic experience is that the testing of the register sources will still take time[10]. This means it is, likely to take more than one census cycle to implement a full register based census.

The steps and processes to assess the register sources, metadata and data are set out in Section 3.

Testing of Processes, Methodologies and IT Systems
Specific processes and statistical methodologies that require testing include statistical register creation, linking and editing. As discussed below, the IT systems that underpin the administrative based census, including databases, data transfer, and data management must be separately tested.

End to End
As with any census, it is important that there is full End-to-End testing. This testing which checks that everything will work together as intended, includes components such as Processing, Analysis and Dissemination.

Specific Testing for Combined Census
Combined Censuses require testing for the Traditional and Administrative components. However, there are specialist tests as well – specifically in relation to the country specific methodology for combining the two main methodologies.

Statistical Testing
There are many different types of statistical testing and each performs different functions. If one type of test is missed, it is critical to ensure that other tests are able to conduct the

---

[10]For more on the Nordic experiences, see UNECE, *Register-based Statistics in the Nordic Countries –Review of best Practices with focus on population statistics*, 2007 -
http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

necessary checks. The main types of statistical testing are shown in Figure 10 and described below.

Figure 10: Testing in Census programme

Type of Testing	Description	Census Methodology	Purpose	Impact on Quality Dimension
Feasibility Testing	Research conducted before the main testing or study. Also known as Proof of Concept.	Traditional Administrative Combined	Answers 'Can this be done?'-	Relevance Accuracy
Cognitive Testing	Small targeted tests	Traditional Combined	Are the questionnaire/ instructions to interviewers interpreted in the intended way?	Relevance Accuracy
Pre-test	A small scale-study	Traditional Administrative Combined	Test and refine applications/proce sses /procedures." Often used to help answer the question "Does it work the way we want?"	Relevance Accuracy Accessibility Coherence
Pilot	A major test of a set or series of processes /procedures.	Traditional Administrative Combined	Answer the question "Do all these pieces work together?"	Timeliness Accuracy Coherence
Dress Rehearsal	Full testing of all activities/ processes/ procedures.	Traditional Administrative Combined	Answers "Will all the parts of the Census work as planned?"	All

Feasibility Testing
Feasibility Testing aims to help answer questions such as "Will this work?", "Can this be done?" This type of testing may include assessing:
- technical feasibility - can the methodology/technique work,
- operational feasibility - can the procedures work, can the methodology be scaled to fit,
- economic feasibility –will the method/approach fit the budget/ provide the necessary savings etc.

Feasibility testing will often be the first test in Traditional, Administrative and Combined Censuses.

Cognitive Testing

Cognitive testing[11] describes methods that aim to capture people's thought processes and understanding in responding to questions and so can help uncover some of the problems people have when answering survey questions. The methods can help to improve questionnaires and interviewer instructions to ensure they are interpreted in the intended way.

Cognitive testing is usually conducted on small, targeted samples, using semi-structured interviews. The testing is usually carried out after the initial design of a questionnaire and before any pre-testing or pilots involving fieldwork. Cognitive testing can also be used iteratively throughout the preparation period to refine questions and minimise response errors. In this way, underlying problems can be resolved before any expensive fieldwork.

As Cognitive testing is conducted on small samples, it can be a useful testing technique to identify problems. However, it does not provide information on the impacts. This type of testing applies to Traditional and Combined Censuses.

Pre-testing

The key purpose is to test whether a specific component (e.g., questionnaire, collection methods, processing system, and data transfer system) is working in the intended way. This type of test will try to copy how the component will work in the main Census and review performance against agreed benchmarks. All types of censuses use Pre-testing.

Pilot Tests

Pilot tests check how the components perform against the benchmarks or KPIs and so build on from pre-tests. They focus on testing the integrated components to answer the question of "Do the pieces work together?" This means that pilot tests need working and well-tested components.

Not all components may be available for a pilot test. For example, a pilot test may focus on testing the integration of fieldwork, internet collection and processing, without any testing of Dissemination. A separate pilot test may focus on testing the integration between Processing and Dissemination.

Examples of pilot tests in an Administrative census include testing the creation of Population and Housing Statistical Registers, integration of linked administrative register data with the Processing system or testing different linking methodologies.

Dress Rehearsal

The Dress Rehearsal is a 'dry run' for the main census, and evaluates all aspects of the census operation. It should be taken under the conditions that are likely to be faced in the main census.

For a Traditional or Combined Census, the Dress Rehearsal includes testing aspects such as the questionnaires, the training of field staff, the effectiveness of the field organization and the

---

[11] For more information on Cognitive Testing, see Scottish Government Social Research Group *Social Science Methods Series Guide 7: Cognitive Testing in Survey Questionnaire Design* http://www.gov.scot/Resource/Doc/175356/0091403.pdf

overall census methodology. Testing of publicity and support systems (e.g. call centres) should also be part of the Dress Rehearsal.

The Dress Rehearsal for an Administrative based Census includes testing all aspects of collecting data from the administrative agencies, preparing the statistical registers and creating the linked census file.

Irrespective of the census method, the Dress Rehearsal should include testing of the Processing, Analysis and Dissemination. There must be time after the Dress Rehearsal to evaluate the findings, and implement necessary changes to systems and processes. Further re-testing may be needed. Information from the Dress Rehearsal will also be used to complete planning of the main census.

For these reasons, it is standard practice to conduct the Dress Rehearsal a year before the Census. This also means that seasonal factors, e.g. weather, holiday patterns, etc. can be replicated. (Note - while weather or seasonal patterns may not affect the NSO in an Administrative census; these factors may affect the agencies providing the register sources.)

The findings of the Dress Rehearsal will help to finalize the final plans, including final calculations of resource requirements. The Dress Rehearsal also helps train staff in preparation for the main Census. This is particularly important for countries who conduct the Census every 10 years. Reporting on the Dress Rehearsal is also provide assurance to the High level Census committee.

The purpose of the Dress Rehearsal is to test readiness and refine budget, resource and time estimates. This means that the Dress Rehearsal is not used to produce statistics. However, it is important to test Dissemination systems and processes.

Specialist Testing

Often specialist tests related to Methodology or Dissemination are also needed. These specialist tests contribute to all the Quality Dimensions.

Methodology testing

All censuses are based on statistical methodologies. The application of these methodologies needs to be fully tested. Areas requiring specialist methodological testing include:

- *Editing and Imputation* – do the techniques work?
- *Coding* – testing coding instructions, code files, coding systems
- *Balancing* – do the techniques work as intended?
- *Statistical Disclosure Control (confidentiality of unit record data)* – do the techniques work?

Dissemination

Depending on user requirements, new or changed output products and systems may be required. These products and systems need to be rigorously tested. It may be necessary to assess the feasibility, as well as testing that the final outputs are correctly produced. This requires specialist testing, as the final tables need to be checked to ensure that the unit record information remains confidential.

Users increasingly require detailed data (e.g. at low levels of geography). The final testing of dissemination products and systems needs to include testing of the final disaggregations.

IT Testing

IT or Software testing is an integral part of the Testing process. It is also an integral part of the process of preparing IT systems and follows the Software Development Life Cycle (also known as Software Development Process) [12]. IT testing has some common testing components.

- **Unit Testing** - where individual units/components of a software system are tested. The purpose is to test or validate that each unit or component is performing as designed
- **Integration Testing** - individual units are combined and tested as a group. The purpose of this level of testing is to expose faults in the interaction between integrated units.
- **System Testing**- where the complete, integrated system/software is tested. The purpose of this test is to evaluate the system's compliance with the specified requirements
- **Acceptance Testing –** testing conducted by the system user. The purpose is to evaluate the system's compliance with the business requirements and assess whether it is acceptable for implementation.
- **Performance Testing** - testing to check the behaviour of the system under different load options

Traditionally this testing is conducted in sequence, with the last stage being Performance testing.

Each of these software-testing components follows a standard approach. The first step is to prepare and agree on a test plan. The test plans show how the software will be tested. The plans may include scenarios, as well as sequences for testing. A series of test cases (mock/test data) will be prepared consistent with the test plan. The final step is to execute the test plan using the test cases. As with statistical testing, Software testing allows for reworking at all stages.

Test Programme

The test programme for the Census needs to bring all the different testing elements together. Through the preparation of an integrated testing programme, it will be possible to understand how the different tests fit together. This helps in planning requirements, identifying test interdependencies, and ensuring the results of each test are used appropriately.

The test programmes will depend on the Census Methodology, as well as the level of risk (including new processes or systems in the census). Figure 11 shows how the different tests can be integrated for a Traditional Census. Figure 12 shows the same testing programme for an Administrative Census. Finally, Figure 13 shows the integrated testing for a Combined Census. In all cases, it is assumed that new processes and systems are being used.

---

[12] See http://softwaretestingfundamentals.com/software-development-life-cycle/ for description of the Software Development Life Cycle

In all of these test programmes, there will be additional work after each test, to refine the relevant component(s). This means that Pre-tests, Pilots, Unit testing, and Integration testing may occur multiple times.

Figure 11: Testing Programme for Traditional Census.

Specialist testing	Main Testing	IT/Software Testing
	Feasibility Testing	Proof of Concept
	Cognitive Testing	
		Unit testing
	Pre-tests	
Editing and Imputation		Integration testing
		System testing
Coding		Acceptance testing
	Pilot test(s)	
		Integration testing
		System testing
Disclosure Control		Acceptance testing
Dissemination	Dress Rehearsal	
		Integration testing
		System testing
		Acceptance testing
		Performance testing
Main Census		

Figure 12: Testing Programme for Administrative Censuses.

Specialist testing	Administrative testing	Main Testing	IT/Software Testing
		Feasibility Testing	Proof of Concept
	Data Quality Assessments		
	Register Creation		
	Linking		Unit testing
	Pre-tests		
Editing and Imputation			Integration testing
			System testing
Coding			Acceptance testing
	Pilot test(s)		
			Integration testing
			System testing
Disclosure Control			Acceptance testing
	Dress Rehearsal		
Dissemination			Integration testing
			System testing
			Acceptance testing
			Performance testing
Main Census			

Specialist testing	Administrative testing	Main Testing	IT/Software Testing
		Feasibility Testing	Proof of Concept
	Data Quality Assessments		
	Register Creation	Cognitive Testing	
	Linking		Unit testing
	Pre-tests		
Editing and Imputation			Integration testing
			System testing
Coding			Acceptance testing
	Pilot test(s)		
			Integration testing
			System testing
Disclosure Control			Acceptance testing
	Dress Rehearsal		
Dissemination			Integration testing
			System testing
			Acceptance testing
			Performance testing
Main Census			

2.4 Ensuring Quality in Census Operational Activities

As noted earlier, a poorly implemented census will have many errors in the various operational activities (**Operational errors**). This will mean that the census will fail to count everyone correctly – **coverage errors.** Censuses with high level of coverage error will have higher non-response/higher levels of missing data (**content error).**

This section focuses on the actions that agencies should take to minimise Operational errors. The overall focus is on the Quality Control actions to correct severe or generally applicable errors, and the Quality Assurance actions to ensure process improvement, with emphasis on the following operational phases:

- o Fieldwork
- o Processing
- o Macro Editing
- o Dissemination
- o Evaluation

The specific issues for Administrative Censuses are discussed in Section 3.

For all of these operational phases, quality is dependent on the following:
- Established and documented processes
- Clear Quality targets

- Monitoring systems
- Well prepared and tested systems, processes and procedures
- Active management support of staff to identify and resolve quality problems

Building quality into Fieldwork

The Fieldwork (Traditional) Census is a complex operation. Members of the public directly complete census questionnaires (on paper and/or internet), or interviewers collect information from respondents. Interview methods include face to face or telephone. Face-to-face interviews can be conducted using a paper questionnaire or handheld devices to capture automatically data during enumeration.

This means that Fieldwork censuses require a large team of field staff (supervisors, interviewers or enumerators) to be recruited, trained and managed throughout the census operations.

The Census Management team, Supervisors and Enumerators need to be focused on the key outcome – a census that correctly counts all the population and their housing units, along with the relevant characteristics. This requires both Quality Control and Quality Assurance.

Quality Control

Before Collection

A number of Quality Control activities should be completed before Collection starts, including:
- Identifying and implementing an appropriate fieldwork management structure
- Check Workload Allocation and Enumeration Areas are of an appropriate size
- Prepare checklists for field staff
- Independently check accuracy of maps
- Test samples of maps, printing or other key resources for compliance against specifications
- Establish procedures and checklists for Supervisors and Enumerators

Training is fundamental. An important role of the Census Management team is to prepare the training material and train staff, to ensure that everyone involved in the fieldwork understands their role in delivering a quality census.

During Collection

Enumerators are expected to follow the collection procedures, and check their own work against provided checklists.

Supervisors should monitor and control the quantity and quality of the fieldwork, in order to meet the required targets. Quality Control tasks include:

- Observing enumerators in sample of visits
- Checking enumerators work against checklists, and assessing the data against provided suitability criteria
- Reviewing workloads for completeness
- Re-interviewing sample of respondents
- Using management information to identify problem enumerators, and taking the relevant actions

The **Census Management (Back Office)** has a key role, including providing management information to assist supervisors perform their duties. However, the back office will also undertake a number of specific quality control actions, including:

- Reviewing sample of supervisor actions, and responding accordingly
- Reviewing workloads for completeness
- Checking that all Enumeration Areas have been covered
- Use management information to identify problem enumerators/supervisors

The key focus is to identify and correct severe errors, such as Coverage errors such as missed enumeration areas or missed housing units, which are best corrected in the field. These types of errors cannot be easily corrected in later stages.

Management Information

Effective Management information systems for Fieldwork are key. They collect information about:
- Pre-Collection activities such as establishment of local census offices, training of field staff, etc.
- Recruitment and training of Field staff, including security checks, training completion and assessment, work completion
- Enumeration progress against targets
- Critical and repeated errors;
- Logistics information such as the shipment of census materials and questionnaires,

Performance should be evaluated against set targets. Possible targets include:
- estimates of the number of housing units
- proportion of occupied and vacant housing units;
- average number of residents per address/housing unit;
- response, refusal and non-contact rates,
- population size,
- population growth rate
- critical error rates
- deadlines for completing work

Historical data, including data from previous censuses, and other relevant data sources such as household surveys and administrative registers can be used to set targets. Significant deviation from the targets may indicate a systematic problem in the collection process. Tools such as dashboards can be used for monitoring and reporting on progress.

After Collection

After Collection, the Field Management team should use the Management Information to review whether all the required Enumeration Areas have been covered, and that targets have been met. Key metadata information, including performance against targets, outliers and exceptions should be recorded. Other relevant information on issues that might affect the data should also be reported.

While it is critical to focus on monitoring and controlling the fieldwork, it is also equally important to consider system or process errors and identify areas for improvement. This is the role of Quality Assurance. All teams have key roles.

Before Collection

As with Quality Control, this starts in the fieldwork preparation stage. The **Census Management** team needs to:

- Ensure Plans are Peer Reviewed
- Establish clear criteria for Staff Selection
- Prepare and deliver Training for all Field staff
- Receive and assess feedback from staff on training
- Test systems and procedures
- Establish and modify Procedures based on testing
- Use previous studies (e.g. other surveys, previous censuses) to identify areas or groups at risk of undercount and being missed in the census.
- Prepare and test targeted strategies to address these
- Establish monitoring and reporting systems, including the management reporting discussed above.

Once the **Supervisors** are in place, it is critical that they understand the different Enumeration Areas under their control. Identifying the areas and/or groups that are at risk of undercount, needs to be a priority. **Supervisors** also need to implement the relevant techniques and strategies to reduce the risk of undercount.

During Collection

During the collection, the focus of Quality Assurance is on preventing errors from reoccurring. Therefore, it is critical to detect errors easily and early, and inform field staff so that they do not continue making them.

While the **Census Management Team** will be busy, it should also identify systematic issues. The following may help to identify systematic problems:-

- Independent Fieldwork Monitoring – a separate team of people to monitor Fieldwork and Field activities
- Procedures and systems to receive and assess feedback from field staff
- Assessments of completed work against pre-determined estimates/targets
- Monitoring the results of Quality Control checks, including monitoring error rates
- Assessment of feedback from field staff, Queries/ Comments from the public and Press and social media coverage (positive and negative)

Supervisors will also identify systematic issues in their areas. Assessing whether the undercount techniques are working as intended will be a priority. **Enumerators** also have an important Quality Assurance role during the Collection. They need to provide feedback on the processes and systems. This feedback is a key way of identifying if there are processes that need to be urgently changed because they are getting in the way of the goal of a quality census.

After Census collection

After the collection is completed, there are many opportunities to identify areas for improvement in future censuses and similar fieldwork activities conducted by the NSO. The focus should be on identifying the effectiveness of the design of the census, including how the processes and procedures worked in practice. These reviews are important to identify the implications for future censuses and other fieldwork. Activities can include:

- Supervisor debriefs with Enumerators
- Debriefs with Supervisors
- Evaluating completed work against pre-determined estimates/ targets, and independent data (e.g. Post Enumeration Survey, administrative data)
- Reviewing expenditure against budget

Feedback from Processing and other phases can also provide valuable information to identify areas for future improvement.

Building Quality in Processing

Processing is key for minimizing Content Error (e.g. non-response and inconsistencies within records) and to ensure that the records used in subsequent processes (Analysis and Dissemination) are plausible and internally coherent. Processing uses a range of statistical methodologies, underpinned by automated systems and data management techniques.

> While Processing ensures that that the records are plausible and internally coherent, it cannot replace high quality enumeration (field or register based). While processing procedures can be sophisticated, if systematic errors occur during collection, the procedures cannot improve the quality of data.

For the purpose of these guidelines, Processing includes all the actions taken on unit records. This includes[13]:

- Receipt and registration of forms/records – including manual or electronic transfer of records to the Processing centre
- Data Capture (if required) – capture of paper forms using key entry, scanning (ICR/OCR) or electronic lodgment of forms
- Coding – assigning of classification codes to responses on the census form or census record.
- Micro–Editing –identifying invalid and/or inconsistent data.
- Imputation - resolving missing, invalid or inconsistent responses
- Balancing – ensuring there is a record for every enumeration area, every household within each enumeration area, and every person within those households.
- Derived Variables – creating new variables using arithmetic formula or aggregation.

Macro-Editing (sometimes called Validation) checks aggregates and combinations of records, and is described below.

---

[13] These steps are based on the steps outlined in the *UN Handbook on Census Management for Population and Housing Censuses, No 83, Revision 1*, 2001, Figure IV.3 Data Processing Cycle in
https://unstats.un.org/unsd/censuskb20/Attachments/Census%20Management-e-GUIDfd30f57d4e0d43f8b4101f1cb3266b7a.pdf

Managing Operational Quality in Processing

The key actions to achieve quality in census processing operations[14] include:

- **Automate** Use as much as possible Automated Coding, Editing and imputation systems and techniques. (Well prepared and tested automated systems produce significantly better quality than manual processes.)
- **Base Edit rules on local Real World situation** Ensure that the edit rules reflect the 'real world', specific to all aspects of the country situation and not a theoretical or ideal model. Otherwise, there is a danger that edit changes will be made just to ensure that the data pass the edits and the final census statistics will not show the real world.
- **Minimise over-editing** Over-editing is the adding of more errors than are corrected. It is important to not edit everything, but instead prioritise edits. All key variables, including variables that are critical for planning, (e.g. sex, age, location) should be edited. However, variables such as disability or literacy work well with less editing.
- **Ensure accuracy of small area data** Regional data at small areas is one of the key products from the census. It is important to ensure that the data for each output area is of an acceptable standard.
- **Build on systems and files used in other projects** Census Processing systems should build on systems and files used in other projects (e.g. Household Surveys)
- **Extensively test** Extensive testing of processing systems, methodologies and procedures must be undertaken, using the type of staff likely to be involved in the operations
- **Implement clear procedures, with appropriate training** Procedures should be clear and up to date, with appropriate training provided to all staff.

Quality Assurance

There are many opportunities to use quality improvement techniques in Processing, as many processes are repetitive and take a reasonable amount of time. It is therefore vital that structures are in place to monitor quality, but also to involve processing staff in identifying problems with quality and in proposing solutions. The following additional guidelines apply:

- Use teams of processing staff to identify and resolve quality problems
- Managers need to ensure that staff comments and observations feed into the quality improvement process. This should be accompanied by appropriate feedback back to staff
- Adopt a continuous improvement approach, including
    - Continually measure the quality
    - Identify the root causes of quality problems
    - Address the root causes of discrepancies
    - Implement corrective action
- Conduct a post-processing evaluation of processing operations, and document the results for future use. Evaluate the processes to identify the lessons learned with the goal of improving each of its components.

Coding

Coding is the assigning of classification codes to responses from the census form or census record. Coding can be automated; conducted by operators using computer assisted coding; completed by operators using manual (clerical) coding; or a combination of all three.

---

[14] This is based on the *Statistics Canada Quality Guidelines for Collection, Capture and Editing,* 2009. See http://www.statcan.gc.ca/pub/12-539-x/2009001/collection-collecte-eng.htm

Quality Control

- Establish an organisation wide central system for all classifications, concordances and code files and use this central system for all census classifications, concordances and code files
- Manual coding operations should use computer assisted coding. Operations should be organised to refer difficult cases to a small number of knowledgeable experts
- Samples of coding work should be regularly checked (through recoding). Checking of complex coding (e.g. occupation coding) may need a higher proportion of records to be checked.

Quality Assurance

- Expert coders should be used to code those cases that are uncompleted after automated coding. These results should be used to improve the reference files.
- Expert coders should also conduct a sample study assessment of the accuracy of the automated coding. These results should also be used to improve the reference files.
- Differences in recoding of computer assisted or manual coding should be shared with the relevant coders to help them improve their coding

Micro-Editing

Micro-editing is the identification of invalid and/or inconsistent data and is conducted on individual records. This is usually a computerized, or computer assisted activity, although some editing operations may be conducted manually.

Overall

- Use *Editing* to identify invalid or inconsistent data and use *Imputation* (discussed below) to resolve missing, invalid or inconsistent responses
- Use *Editing* and *Imputation* to eliminate the most obvious inconsistencies
- Make the fewest required changes to the originally recorded data

Quality Control

- Monitor the editing work, including frequency and volume of edit failures. Monitoring should be conducted at appropriate disaggregations (e.g. region, key/non key variables[15], collection mode, and language of the collection).
- Where manual editing is conducted, use expert editing staff to edit independently a sample of records and compare the results with the original sample.
- Reapply edits to units that have been corrected to ensure that no further errors are introduced directly or indirectly by the edit correction process.

Quality Assurance

- Ensure that all edits are internally consistent
- Use Macro-Editing to provide feedback on the quality of editing of key variables

Manual editing and coding – special issues

- It is important to check a sample of manual editing and coding, including computer assisted editing and coding. This checking aims to ensure that staff follow the <u>correct processes.</u>

---

[15] Key variables are the most important variables in the Census – these include Location, Age, Sex and Nationality.

Another processing staff member should re-edit or re-code the records. The results should be compared to identify systematic patterns.

- Check the work of all staff. High proportions should be checked at the start and the end.
- Experts should be used for difficult technical decisions.

Imputation

Imputation is the resolving of missing, invalid or inconsistent responses to ensure that the data is plausible and internally coherent. Using specified procedures, wrong or missing values are corrected by using other data items within the record or from records of other households or persons.

There are two main computerized imputation techniques. Static imputation or "*cold deck*" is mainly used for missing or unknown items. Dynamic or "*hot deck*" imputation is used for inconsistent or invalid items as well as missing data. Both methods are based on using values from a "donor" that has complete observations for all variables and similar characteristics to the incomplete or incorrect observation (recipient record). The donor records are stored in imputation matrices, similarly to a pack of cards.

The international best practice is to adopt integrated Editing and Imputation software. The most common Census tool is CANCEIS [16] prepared by Statistics Canada. This brings editing and imputation together in one automated system. Through using this type of automated rules based software, the number of staff needed for editing can be minimized (or even completely removed).

In addition, there are manual methods – assigning wrong or missing values to Not Stated, Manually or computer assisted correction of values that are wrong or missing

However, irrespective of the methodology, all require clear and tested rules. The solution that is used must be repeatable, consistent and produce correct values.

Cold-Deck and Hot-Deck Imputation

In *cold deck imputation*, the donor records come from a predetermined set of records, which are used throughout the census. (The records may come from 'clean' census records or from other sources.) This set of records are not changed after processing the records for the first, second, tenth or any other persons.

Hot-deck Imputation uses actual responses provided by other census respondents, which are continuously updated as records are edited and imputed. The donor records constantly change as records are updated and/or by logically "shuffling the deck". This gives the term "hot deck". The values stored in the hot deck represent information about the "nearest neighbours" with similar information. For example, when a person's marital status is unknown, the hot deck will contain information about the most recent person encountered with the same sex, age, living arrangements and valid marital status.

---

[16] See Guertin (2014) *Editing the 2011 Census data with CANCEIS and options considered for 2016*. Working paper, UNECE Work Session on Statistical Data Editing, 2014
https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2014/mtg1/Topic_4_Canada_Guertin.pdf

When the editing system finds an acceptable value for a data item, it puts it into the imputation matrix. When it finds an unacceptable one, imputation replaces it with a valid value from the imputation matrix.

Computerized Imputation packages and the programs within those packages use cold deck and hot deck imputation in different ways.

Quality Management in Imputation

While Imputation, especially together with Editing, can speed up the process of identifying and correcting errors, it is still important to manage the quality. Key quality steps include:

- **Monitor the levels of missing or inconsistent responses**. When the percentage of missing or inconsistent responses is high (5 to 10 per cent, or more, depending on the situation), imputation may distort the census results.
- **Use Audit trails, performance measures, and diagnostic statistics** to analyse the quality of the edits and the speed of processing
- **Preserve as much respondent data as possible.** The imputed record should closely resemble the failed edit record. Imputing a minimum number of variables is usually best
- **Edit the imputed record -** The imputed record should satisfy all edits
- **Flag imputed values.** The methods and sources of imputation be clearly identified
- **Retain un-imputed and imputed values** to evaluate the degree and effects of imputation.
- **Run the Editing and Imputation together, multiple times.** For example, a first run is conducted to perform the actual editing to identify error records. Where hot-deck imputation is used, this run will also update the imputation matrix. The final run makes certain that no errors remain in the dataset and that the editing program did not introduce new errors.[17]

Balancing and Creating Derived Variables

Balancing

Balancing ensures that there is a record for every enumeration area, every housing unit and every known person within those housing units. In this way, the census balances across the different units.

Balancing is a key element of Quality Control in Census Processing for all Census methodologies, including Fieldwork, Administrative censuses and Combined Censuses. If records have been missed during previous steps, the balancing step will identify what is missing, and make appropriate corrections.

Corrections are made by adding records where housing units or individuals are known to exist, but where records cannot be located in the census database. Records are only added where there is clear evidence. These additional records are commonly called "*Substitute or Dummy Records*". They are always clearly flagged as Substitute or Dummy records.

Substitute or dummy records are created in the following situations:

---

[17] The *Handbook on Population and Housing Census Editing* Revision 1, United Nations, 2010, https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1e.pdf, contains more information.

- **Housing unit missing at least one known individual**. Where the dwelling record and at least one individual record was received, substitute individual records will be created for the missing individual record(s)
- **Occupied Address/Housing Units with no dwelling record, but individual record(s).** A substitute dwelling record will be created.
- **Dwelling known to be occupied, but no dwelling or individual records received**. In this case, a dwelling record and an agreed number of individual records will be created

The first step is to establish a set of rules for the creation of substitute records. These rules should reflect the country specific situation. The next step is to check there are individual and dwelling records for all occupied dwellings. This includes linking records for individuals with the correct address/housing unit record. Similar checks are done to ensure that all individuals can be linked to a valid occupied address/housing unit record.

Separate checks are conducted to ensure there are records for every address/housing unit in each enumeration area. (For Administrative censuses, the enumeration area may be the smallest level of administrative geography – e.g. village, willayat, suburb, etc.).

If the checking identifies gaps, then it will be necessary to create substitute records. These records may also have values imputed for key variables (e.g. location, sex, age, nationality), but other variables will be left as not specified. These records will also be flagged as substitute records.[18] In most cases, the Imputation system described above will be used to impute these key variables.

Substitute forms as an indicator of Quality

The number of substitute forms provides a measure of the quality of the overall collection and the management of records within processing. Internationally, the number of substitute forms is increasing, as the environment for collection becomes harder.

Creating Derived Variables

In order to get the best use out of census data, countries often need variables that are combinations and variations of other variables. For example, Age should be determined by collecting Date of Birth and then subtracting the date of birth from the census reference date. This information would be stored on the record.

Rather than having to develop a program to recode the information each time it is required, these new variables can be created once and then stored on the individual or address/dwelling unit records. These new variables can be used for Analysis and Dissemination in standard or specialist outputs. This process of creating these new variables is commonly called "Creating Derived Variables.[19]"

---

[18] For examples of how countries create dummy or substitute records, see *Understanding substitution and imputation in the 2013 Census,* 2014 – Statistics New Zealand http://www.stats.govt.nz/Census/2013-census/methodology/substitution-and-imputation.aspx and *Item Edit and Imputation: Evaluation report June 2012,* Office of National Statistics http://webarchive.nationalarchives.gov.uk/20160107185728/http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/index.html, 2012

[19]See also Handbook on Census Editing Revision 1 for more information

National statistical/census offices need to decide what standard new variables are required and what might just be specific to particular outputs. The following should be addressed in order to manage quality of Derived Variables:

- **Create Derived Variables only once** - If a new variable will be used more than once in Analysis or Dissemination, it should be created once as a standard derived variable.
- **Carefully specify and test the creation of derived variables** – Specialist subject matter statisticians may need to specify and test that the derived variables are created correctly
- **Remember to edit the Derived Variables** –Specific edits should be used to ensure that the Derived Variables are consistently calculated.
- **Managing expectations of Quality of Derived Variables** – Derived Variables are based on combinations of other variables; this means that the accuracy (e.g. non-response rates) of the Derived Variables may be less than the component variables.

Macro Editing – Key to ensuring that final data meets requirements

Macro Editing (also known as Validation or Analysis) checks that the final overall data meets agreed minimum standards. Macro Editing is conducted on aggregated data. It focuses on checking aggregates (totals and sub-totals), and distributions of edited data against predicted frequencies and tolerances to identify any remaining problems with the data.

Because errors are not always obvious at the unit record level, it is really important to undertake checks at the macro level. This includes checks of

- Aggregates – e.g. totals and sub-totals
- Ratios – e.g. Sex ratios, unemployment rates
- Derived Variables – e.g. Labour Force Status
- Distributions
- Outliers

Macro Editing compares census data with other sources as well as comparing trends over time. Macro editing improves the coherence of the census statistics, helps the analysts understand the data and so increases user confidence. Macro editing also has a number of Quality Assurance functions, as it provides direct feedback on the Processing functions and to the Census and other survey designers for the future. (Note if errors found in Macro editing affect individual records, it is best to return the records back to Micro-Editing for more processing.)

Macro Editing Techniques

The techniques are designed to identify suspicious values and inconsistencies in distributions and aggregates. These techniques include

- Internal Consistency Checks (such as Reasonableness, Demographic techniques such as Age-Sex Pyramids, Sex ratios)
- Comparisons with other statistics (including published and unpublished statistics)
- Assessing consistency with Metadata created in the earlier phases

These techniques are briefly discussed below. Many of the techniques may also be used in the Evaluation phase. More information will also be provided in the forthcoming GCC Census Evaluation Guidelines.

Internal Consistency Checks

Reasonableness Checks

The purpose of these checks is to determine whether the census statistics match with expectations or benchmarks. Expectations will be based on other statistical information (demographic, economic, social statistics), as well as demographic methods. For example, demographic techniques can be used to calculate expected estimated counts of the population, which are then used as benchmarks for the census.

If the statistics look suspicious, consider the following
- Does the metadata explain any unexpected differences?
- Are the statistics are consistent with other sources?
- Are changes shown in the statistics consistent with real world changes?

Demographic Techniques

Demographic Techniques such as Age-Sex Pyramids and Sex Ratios can be used to assess whether the census statistics are coherent with other population statistics.

The **Age-Sex Pyramid** is a graphical representation of a population's age and sex structure at a point in time. Analysis of Age Sex Pyramids can identify possible inconsistencies, including missing data for specific age groups. More detailed analysis, e.g. of single year age groups may then be needed[20].

Sex ratios refer to the ratio of Males to Females. The standard sex ratio at birth is between 102 and 107 – that is there are between 2 and 7 % more boy babies born than girls. The sex ratio generally declines with age. In most cases, the older populations (e.g. 75 +) will have more women than men.

Comparing the sex ratio from the Census data, with these standard ratios helps to identify if parts of the population have been missed. It is important to conduct the comparisons using standard tools, but in the context of local situations[21].

Comparisons with other sources

Other sources of statistics can be used to assess census aggregations and distributions. These sources include Survey sources, Administrative statistics and previous census results. Possible sources are shown in Figure 14

---

[20] While Demographic theory contains standard Population pyramids, it is important to remember that that the Age-Sex Pyramids in GCC countries differ to the standard models.

[21] For example, sex ratios for the non-citizen populations are very different to the standard pattern, as most are male.

Figure 14: Possible Sources for comparing Census statistics

Survey sources	Labour Force Survey, Demographic and Health Survey (DHS) Multiple Indicator Cluster Survey (MICS) Household Income and Expenditure Survey (HIES) Economic statistics
Administrative based Statistics	Vitals (Births, Deaths) Marriage and Divorces Health Education
Previous Census results	

Significant differences should be investigated. In all cases, explanations should be prepared.

Consistency with Metadata

Metadata provides important contextual and explanatory information, which can help understand differences in the data versus expectations. In a traditional fieldwork census, metadata from the Collection phase and Processing may help explain differences. Administrative censuses will have a large range of metadata, including quality information about the Register Sources. Processing metadata such as error rates or imputation rates can also explain differences.

Dissemination

The role of the Dissemination phase in any census is to
- Deliver relevant products and services
- Maintain accuracy of the data
- Provide timely statistics in a predictable schedule

It is also important to see Census Dissemination as part of the wider Dissemination processes of the National Statistical Office, which will serve the users over a long period of time, rather than something that is specific only to the Census project.

Managing Quality in Dissemination

All of the Quality Dimensions apply in the Dissemination phase.[22] .

Relevance

Relevance of the census can only be achieved by producing and delivering relevant products and services. Therefore, it is necessary to:
- Review user experiences of previous census outputs/products
- Consult current and potential users on requirements for products and services, including types of products and services, types/levels of data disaggregation, timeframes, etc.
- Align Products and services with user requirements and expectations

---

[22] While many aspects are under the control of the team responsible for Dissemination, some elements such as meeting the overall project timetable depend on other census activities.

- Ensure that Disclosure Control routines (confidentiality) are correct, clear and well documented

Accuracy

Providing accurate products and services requires:
- Extensive testing of products, services and Disclosure Control routines
- Consistently following Quality control checks (see below)
- Preparing and providing appropriate Metadata for users, including a data dictionary describing all the census variables, definitions, classifications etc.

Timeliness

Timely census data requires:
- Census Release calendar, published early in the census cycle.
- Ensuring that the published calendar is reflected in the project timetable
- Following the agreed timetable for preparation and testing of Dissemination tools, systems and services
- Following overall project timetable for collection, processing and analysis of census statistics

Accessibility and Clarity

Delivering accessible and clear outputs requires that
- Outputs are provided in formats that meet user requirements
- Metadata is provided to help explain the data
- A Team exists to support data users, including providing training in the products and services and on call support

Coherence

Coherent census outputs are:
- Published using agreed international or regional standards
- Compared with other statistics and any differences shown in the metadata

Quality Control Checks in Dissemination[23],

Quality Control of Census products and services is very important. In addition to the checks set out in the NSO Dissemination policies, additional Quality Control checks could include:

Checks of Tables and Other Data Products.
- Checklists for reviewing tables and other data products should include:
  - Checks that totals and key aggregates match control totals
  - Tests of all links in electronic products
  - Checks that Disclosure Control techniques have been used correctly
  - Checks that all the relevant metadata is provided and that it is accurate
  - Checks that the data and text are consistent in both languages.
- Peer reviewing all products before release. The peer review should include assessing
  - Soundness of the data,
  - That the data and text are consistent in both languages.

---

[23] Based on *Statistics Canada Quality Guidelines- Data Dissemination and Communication* , 2009
http://www.statcan.gc.ca/pub/12-539-x/2009001/dissemination-diffusion-eng.htm

o Appropriateness for publication

- Additional checks for reviewing written reports and publications should include:
  - o Thoroughly double-check numbers, reference periods (e.g. "in the last six months" and "compared to last quarter") and words that depict trends (e.g. "increase" and "drop").
  - o Avoid repeating numbers provided in tables in the text; otherwise make sure they are the same.
  - o Verify numbers in articles and publications against those provided in other products, including in data portals.
- Peer reviewing all products before release. The peer review should include assessing
  - o Soundness of the data,
  - o Soundness of the analysis, including consistency with other statistical reports published by the NSO
  - o That the data and text are consistent in both languages.
  - o Appropriateness for publication

Quality Assurance in Dissemination

Lessons from the Census Dissemination phase can be very useful for other dissemination activities. Therefore, it is important to:

- Ensure that staff comments and observations are fed into the quality improvement process.
- Adopt a continuous improvement approach, including
  - o Continually measure the quality
  - o Identify the root causes of quality problems
  - o Address the root causes of discrepancies
  - o Implement corrective action
- Conduct a post-processing evaluation of Dissemination operations, and document the results for future use. The focus should be on identifying the effectiveness of the Dissemination of the census, including how the processes and procedures worked in practice.

Reporting on Quality of Disseminated Census products

Regularly report on the quality of the Disseminated Census products. This includes reporting on:

- Availability of products at different levels of detail, formats and media and Frequency and level of product use. These provide indicators of **Relevance**
- Time lag from the Census reference date to the release of the product. This is an indicator of **Timeliness** with respect to users' needs.
- Time lag between scheduled release date and actual release date. This is a measure of **Punctuality**.
- Occurrence of errors detected after release. This is an indicator of **Accuracy** and/or **Coherence**
- Feedback from users on timeliness, accessibility, availability and perceived accuracy of the final statistics can help measure all the Quality Dimensions.

Evaluation

A standard phase in the Census project is the formal evaluation of the overall quality of the results from a <u>statistical</u> perspective. This may be done by conducting a post enumeration survey (to measure coverage and content errors), comparing the census results with similar data from other sources and by using demographic techniques and analysis.

The purposes of evaluating the accuracy of the data are to inform users of the quality of the current census data and to assist in future improvements. Future improvement may be achieved by:

(a) Improving processes, and

(b) Establishing performance benchmarks against which the quality of the data from subsequent censuses can be measured.

Evaluation of data accuracy has two parts. Preliminary evaluation will enable the identification of any problem areas that have not been detected in the earlier quality management processes. More evaluation should be undertaken on data items where problems have been identified or where new questions or processes have been used.

Evaluation is a key aspect of the overall Quality Assurance of the Census, as it provides an overall assessment of the overall quality of the Census statistics.

Additional guidelines related to the Evaluation of the Census are expected to be published by GCC-Stat in early 2019.

3 Specific Quality Considerations for Administrative Censuses

Administrative data refers to information collected for administrative reasons. Government agencies and other organisations collect this information as part of the process of providing services to businesses, citizens, residents and other clients and customers. The records may take the form of registration of customer/client information as well as transactions.

Administrative data is often used for operational purposes and the statistical use is secondary. This may mean that the NSO needs to transform the administrative register to meet statistical requirements.

This transformation is particularly important for Administrative based censuses, which bring together many different administrative sources. These sources have varying administrative purposes and so have different definitions, reporting frequency, etc.

This means that the Administrative Census will need to manage a number of additional quality issues, including:
- Limited or lack of quality control over the data - the NSO does not have any direct control over the data or the administrative processes used to collect it
- Possibility of having missing items or missing records (an incomplete file)
- Differences in concepts – leading to bias and coverage problems
- Timeliness of the data (it is possible that due to external events, some or all of the data may not be received on time)
- Need for NSO to invest in systems and expertise to clean and combine the data to produce census statistics.

> In summary, the major difference is that the overall quality of an administrative census depends on the quality of the Administrative registers, owned and operated by administrative agencies and the management of the register data within the NSO.
>
> This means that Administrative censuses need additional quality processes and checks.

3.1 Quality Assessment of Administrative Data – General Principles

Quality Assessment of Administrative data is an ongoing and iterative process of assessing the data's fitness for statistical purposes. [24] It covers the entire statistical process, and involves monitoring data over time, and reporting on the variations in the quality. This includes many of the Quality Control and Quality Assurance checks set out in Section 1.

However, these quality methods are of limited value if the underlying administrative data are of poor quality. This means that NSOs need to investigate the administrative data to
- understand the data;

---

[24] This section draws on the UK Statistics Authority *Quality Assurance of Administrative data regulatory standard, Quality Assurance of Administrative Data – Setting the Standard*, 2015 https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-settingthestandar_tcm97-44370.pdf

- identify any errors, uncertainty or bias in the data;
- make efforts to understand why the errors occur and to manage them,
- determine whether the administrative data is suitable for statistical purposes and
- communicate to users how the use of administrative data could affect the statistics and their use.

If the data contains errors, uncertainty or bias; then the NSO needs to carry out the following actions:
- evaluate the likely impact on the final statistics,
- establish whether the data can be resolved, or whether there are actions that can be taken to mitigate the risks (e.g. only use part of the data, use alternative sources); and
- determine whether the users need to be notified.

Often the investigation and resolution of these issues is complex and takes time, staff and financial resources.

An important part of the quality assurance of administrative data is the working arrangements and relationships with the administrative agencies providing the data. In some cases, statistics may only rely on transfers from a single data supplier. In other cases, such as the Census, there may be multiple suppliers.

Data may be provided directly from the organisation that records the data, or via an intermediate organisation. If the data is provided directly from the administrative agency that records the data, the NSO should engage directly with the agency to understand their quality assurance and quality control actions.

If the data is provided by an intermediary organisation, e.g. by a Ministry who collects the data from regional or local agencies, then it is important for the NSO to understand the full data cycle, including understanding the quality assurance processes carried out by the original data suppliers and the intermediate organisations.

Administrative data used in official statistics may also be subject to different kinds of audit – for example, financial or procedural audits. These audit investigations may also be useful for providing information on the data quality.

In summary, the Quality Assurance of Administrative data should include:

- The Operational context for the administrative data collection
- Communication with the agencies supplying the data
- Quality Assurance principles, standards and checks conducted by the agencies providing the data
- Quality Assurance activities undertaken by the NSO, including the compiling and publication of relevant documentation for users.

NSOs who are making extensive use of administrative data for statistical purposes will be conducting regular assessments of the administrative sources, using this type of framework. The Census project should use these assessments and share findings with other parts of the NSO. This then provides an integrated approach to understanding and improving the quality of administrative data.

3.2 Quality of Register Sources

It is therefore vital that NSOs determine the quality of the register sources. The framework developed by Statistics Netherlands, known as the Hyperdimension model[25], recognises that the quality of administrative register data occurs at three levels (Hyperdimensions) - Source, Metadata and Data.

The **Source** Hyperdimension relates to where the data is extracted. This includes the register as a whole, the register owner (the agency responsible for the register) and the environment for the register. **Metadata** means the information about the items in the register, including definitions and classifications. This also includes information about actions undertaken by the register owner to treat the data (e.g. update, correct or change the data). **Data** means the observed facts in the data.

Figure 15 shows how these Hyperdimensions relate to the Quality Dimensions. As the Figure shows, many of the dimensions, including Relevance, Timeliness and Punctuality are determined in the Source Hyperdimension. The Accuracy quality dimension is determined in the Data Hyperdimension. Issues of Accessibility and Coherence relate to all three Hyperdimensions.

Figure 15: Quality Dimensions and Hyperdimensions of Administrative data

Quality Dimension	Hyperdimension		
	Source	Metadata	Data
Relevance	■		
Accuracy			■
Timeliness and Punctuality	■		
Accessibility and Clarity	■	■	■
Coherence and Comparability	■	■	■
	High Level		Detailed

Key

■	Quality Dimension relevant to this Hyperdimension
	Quality Dimension not relevant to this Hyperdimension

The Hyperdimension approach enables High Level and Detailed issues to be addressed in a staged manner. High-level issues (found in **Source** and **Metadata**) are addressed first. When these issues are addressed, detailed investigations requiring in-depth analysis of **Data** are conducted

---

[25] For an overview of the Hyperdimension model, see Piet Daas, et al, *Checklist-quality-evaluation-administrative-data-sources,* 2009 http://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60

The Quality Dimensions discussed in Section 2 also apply to Administrative Censuses. However, as Figure 16 shows, Administrative Censuses have extra quality considerations. The Hyperdimension model enables these considerations to be addressed in a systematic way.

Figure 16: Additional Quality Considerations in Administrative Census

Quality Dimension	Examples of Additional Quality Considerations
Relevance	Are all the required topics (data basket variables) available in the registers?
	Do the definitions and classifications used in the registers meet requirements?
	Do the final Population and Address Registers cover all the country?
Accuracy and Reliability	Is the coverage of the population acceptable?
	Is each unit in the registers assigned a geographic code?
Timeliness and Punctuality	Are registers available for the census reference date?
	Are the registers updated on a timely basis?
	Are the registers available in a timely manner?
Accessibility and Clarity	Are registers available in a standard format?
	Are the registers available and accessible to the NSO?
	Is there clear information (metadata) about each of the registers?
	Can the registers be linked together?
Coherence and Compatibility	Are the register sources internally coherent?
	Is the linked data from the registers coherent?
	Do the registers use international or regional standards?
	Does the metadata meet statistical requirements?

3.3 Environment for an Administrative Census

In addition to availability of register data, there are a number of pre-conditions[26] which must be in place before an Administrative based census can be implemented. These pre-conditions are:

- Appropriate Legislation
- Strategic and Political Support
- User Requirements
- Availability of Administrative Registers
- Statistical Use of Registers
- Quality Frameworks
- Methodology
- Infrastructure
- Plan
- Financial and Human Resources
- Organisation of Census

---

[26]See *Pre-conditions for Administrative Register Census,* GCC-Stat, 2015, http://gccstat.org/ar/elibrary/publications/gccstat/item/gcc-pre-conditions-for-an-administrative-census-in-gcc-countries-2

The success of any Administrative Register requires these pre-conditions to be in place. Without them, there will be a range of quality issues, with significant impacts on the census statistics, as can be seen in Figure 17.

Figure 17: Impacts of Pre-Conditions not in place

Pre-Condition	Related Quality Issues	Examples of Impact
Appropriate Legislation	Legal environment limits access to any unit record administrative data	Not able to conduct Administrative based Census
	Legal environment restricts availability of data for all topics	Final Outputs not **Relevant** to users
Strategic and Political Support	Lack of strategic support restricts availability of data for all topics	Final Outputs not **Relevant** to users
User Requirements	User Requirements not clearly identified	Final Outputs not **Relevant** to users
	Products and Services prepared with minimal input from users	Final Outputs not **Relevant** to users, not **Accessible** in required ways
Availability of Administrative Registers	Core registers (Population and Address) don't exist / can't be created	Not able to conduct Administrative Census
	Registers for other topics don't exist	Not able to provide full range of topics, so Final Outputs are not **Relevant** to users
Statistical Use of Registers	Registers don't contain unique records	Outputs don't represent country and so are not **Accurate**
Quality Frameworks	Frameworks for assessing quality unclear or inconsistent	Outputs not **Accurate.** Unable to assess **Coherence**
Methodology	No Methods for Linking	Not able to conduct Administrative Census
Infrastructure	Transfer methods not in place, NSO can't receive administrative data	Not able to conduct Administrative Census
Plan	Tests not scheduled	**Accuracy** and **Coherence** of data impacted
	Plan does not cover Dissemination.	Release of outputs not **Timely**
Financial and Human Resources	Resources not Assigned	Outputs are not **Accurate**, **Timely** or **Coherent**
	Skills not Available	
Organisation of Census	Lack of High Level Committee means no forum to resolve strategic quality issues	Outputs are not **Accurate**, **Timely** or **Coherent**

As can be seen, without some core pre-conditions in place (e.g. Administrative Registers, appropriate Infrastructure to transfer and receive the data), it is not feasible to conduct any form of administrative census. Without the other pre-conditions in place, the resulting data will not be of appropriate quality. .

It takes time to put all these pre-conditions in place. Therefore, it is critical that the NSO assesses the pre-conditions, identifies how to fill any gaps and implements a clear plan to fill the gaps. It is also helpful to conduct regular assessments to ensure all pre-conditions remain in place.

3.4 NSO Processes

An Administrative Census requires in-depth investigation, analysis and transformation of the register data to meet the census statistical requirements. As the NSO has no direct control over many or most of the administrative sources, it needs to put good processes and procedures in place to manage and properly analyse the register sources.

This means that Administrative Censuses require considerable investment in the Preparation phase. Once the NSO is confident that the Pre-Conditions for an Administrative Census are in place, then an extensive range of technical work needs to be undertaken[27]. This includes:

- Identifying Possible Data Sources
- Detailed Assessment of the Administrative registers[28]
- Designing the Census, including the mix of Administrative and any Fieldwork
- Identifying and testing the Data Manipulation Methodologies
- Preparing IT requirements
- Designing and Building IT systems and procedures to manage the full range of Administrative records
- Creating and testing trial Statistical Registers

These technical activities may take place over a long period of time. Many iterations of the trial Registers may be needed, to ensure that the statistics meet quality standards.

3.5 Assessing Administrative Records - Quality Checklists

Administrative data needs specific quality assurance checks. In a Traditional Census, the NSO can control quality, including through the specification of requirements, testing of questionnaires, management of the collection and processing. However, with administrative records, the NSO does not have this control.

Administrative based censuses bring together many different administrative sources. Therefore, it is critical that the NSO conducts a thorough assessment of the quality of all the administrative records.

To help NSOs in this task, GCC-Stat has prepared a set of Administrative Records Quality Assessment Checklists (See Appendix 1). The checklists will be used to determine what register sources and data items to use in the Administrative Census. They will also be used to prepare Metadata and to prepare the Quality Reports described in Section 4.

This section provides an overview of the checklists.

---

[27] The *Guideline to Planning and Preparing Administrative Register based censuses in GCC Countries,* GCC-Stat; 2018 describes all the steps for planning of Administrative Censuses in the GCC.
[28] The Checklist in Appendix 1 can be used as the basis for this assessment

There are five checklists:

1. **Source Checklist** – checks of **each potential administrative register source**.
2. **Metadata Checklist** – checks of metadata **for each suitable source**
3. **Data Checklist** - detailed checks on the **data**, including population units, identifiers, data items and their respective values
4. **Linking Checklist** –reviews whether the **linking methods** have worked as planned.
5. **Statistical Dataset Checklist** – final check by the NSO that the **final administrative based dataset meets the census requirements**.

The checklists include Review and Decision Points to help ensure a quality census dataset.

Checks are first made about each possible register, using **Checklist 1.** Subject to the results, the **Metadata** is reviewed using **Checklist 2.** Where the Metadata is suitable, the detailed **Data Checks** set out in **Checklist 3** are completed. These identify the specific Administrative Register data items and population units that can be used.

Once all the identified administrative registers (and their associated metadata and data) have been assessed, the NSO can determine the technical feasibility of conducting an administrative register based census. This includes identifying the specific population units and data items that will be used from each administrative sources.

At that point, the register data should be brought together. A further set of technical checks should be conducted of the **Linking Processes**, set out in **Checklist 4** as well as a final separate check on the overall Integrated **Statistical Dataset** using **Checklist 5.**

Instructions for completion

Each checklist covers several quality areas, containing a series of indicators, scored by filling one or more questions. In some cases, the response is descriptive (e.g. describing the register owner's processes). In other cases, the response is numerical.

While most information is obtained from the Register Owner (the agency responsible for the register), some information also needs to be obtained directly from the NSO. A number of checks also involve comparisons of information provided by the Register Owner with the requirements of the NSO.

Most checks in the Data Checklist include an indicator. These can help determine the relative quality of data items from different sources.

3.6 Managing Identifiers

There are specific quality issues for the Identifiers used to link registers.

Countries in the GCC have identity systems. One administrative authority is responsible for the management of these government identity numbers (ID Nos) in different government registers. In this case, the relevant authority is responsible for the overall integrity of the government ID numbers. The individual administrative agencies are responsible for their implementation in agency processes, including ensuring they are using the up to date ID Numbers.

To protect confidentiality, NSOs will typically replace the Government ID Number with a specially created Statistical ID Number. For example in the case of the Population Register, a small team of people in the responsible agency (or in some cases the NSO) will replace the Government ID with a Statistical ID Number. Only the Statistical ID Number will then be used for statistical purposes.

While all countries in the GCC have government ID systems, in some cases they are not fully implemented in all administrative processes. In some countries, not everyone will have a standard Government ID number[29]. This may mean that some registers to be used in the Census may not have Government ID numbers for everyone. Some registers may not have any Government ID Numbers.

In these cases, it will be necessary to link records based on a range of data variables. These variables may include date of birth, name, mother's date of birth, mobile phone number, address (where available). This form of linking uses probabilistic matching techniques. They create a statistical identifier based on these variables, and so enable the registers to be linked.

The number and choice of variables are very important issues when deciding how the registers will be linked.

Creating Statistical Identifiers

If a statistical identifier is required, then all the variables to be used in creating this identifier need to be edited.

- Addresses should be coded to a standardized format. This can include coding to a unique value such as GPS code or to a standardized address format
- Any spelling errors in text fields (e.g. name) should be reduced.
- Creating lists of synonyms for names - While Arabic spellings of names may have less variability; there may be differences in spelling of names recorded in English. The variety of nationalities in the GCC mean a variety of names, and many different spellings of similar names. Even small differences in names can have an impact on the likelihood of being classified as a match[30]. It is recommended that lists of names with common aliases (including nicknames) be prepared.
- Similar lists, with synonyms, can be prepared for occupation. For example, a list for occupation might include Teacher, Professor, Lecturer, Instructor linked as similar[31].

As situations differ between countries, this means that local solutions will need to be prepared. For example, in some GCC countries all names are always stored in all registers in both Arabic and English. However, some countries may have registers with names only recorded in one language (e.g., some Education registers may only contain records in English).

---

[29]For example, practices for issue and use of Government ID numbers to children varies across GCC countries.

[30] For example, there are many versions of the name Mohammed recorded in English. Possible other spellings include Moohammed, Mahmad, Mehmed, Mahamed, Mohamad, Mohamed, Mohammad, Muhammed, Muhamad, Muhamed, Muhamet, Muhammed, Muhammet, Mahammud, Mehmet, Mohd, Muh,"Mohamed",and "Mahamid".

[31] See Anders Wallgren and Britt Wallgren *Register-base Statistics. Statistical Methods for Administrative Data*, Second Edition, 2014, for more recommendations on standardisation.

There are a number of different Data management techniques to help improve the quality of these linking variables. One technique is **Parsing** – dividing a string of text into separate variables. For example, Date of Birth can be split into separate variables for Day, Month and Year. In this way, one complex variable is divided into three variables that can be treated separately. The effects of typing errors and variations can then be reduced. Accuracy of linking is improved as the rates of false match and non-match reduced.

Another useful technique is **blocking**. One or two variables can be used to divide each register into a number of smaller registers for each category of the blocking variables. An example of a blocking variable could be City/Governate, or Citizen/Non-Citizen.

Types of Linking errors

Errors in linking can occur because of errors in the linking variables (Statistical identifiers) or because of issues with the Statistical Units. Typical errors include:

- Errors in the Linking variables – can occur when the Statistical Id Number comes from multiple variables.
- Changes in the Linking key over time – an issue when registers from different time periods (with different practices with ID numbers) are being compared.
- Linking keys are correct – but statistical units are wrong. A common census example is multiple census households with the same physical address. If households share the same physical address (e.g. in a traditional Arab house or a large housing complex), the Government ID systems may show the same physical address for many households. In this case, the Linking keys (ID number, ID number for address) will be correct, but each of the households need to be separately identified.
- Statistical Units have changed between reporting periods or reference dates. For example, if a Population Register and register of Marriages and Divorces relate to slightly different time periods or have different updating schedules, then an event such as a marriage may not be shown in both. This means that the family may be recorded differently in the two registers.

4 Measuring and Reporting on Quality

The focus of the previous sections has been on the Quality Control and Quality Assurance actions to manage quality. However, a key element of quality management is measuring and reporting on Quality.

GCC-Stat recommends that standard quality reports be published in conjunction with all الهيكل والتوصيات المتعلقة بإعداد تقرير الجودة لإحصاءات دول مجلس التعاون لدول published statistics. See الخليج العربية (Recommendations for preparing quality assessment report of GCC statistics)[32]. These reports allows users to be informed about the limits and constraints. GCC-Stat has prepared a standard reporting template for regular statistics (e.g. quarterly, annual, etc.). This reporting template provides the users with information across all the Quality Dimensions in the GCC Data quality framework described in Section 1. A Census specific reporting template, based on the GCC standard reporting template has been prepared and is shown in Appendix 2.

4.1 GCC Census Quality Reports

The 2020 Census Quality Report should include the following:-

Introduction

This section should include:
- A brief history of the Census, its methodology and approach, and main outputs
- Reference to other documentation (questionnaire (where relevant), methodology, Metadata such as Data Dictionary, etc.)

Assessment of Quality Dimensions

Each dimension should be assessed in turn.
- Relevance.
- Accuracy
- Timeliness and Punctuality
- Accessibility and clarity
- Comparability and coherence of data.

User Needs and Perceptions

This section should include a short description of the main uses of census data, the main users, and their feedback.

Conclusion

A summary conclusion should include
- Main quality problems encountered in the 2020 Census
- Recommendations for improvement
- Follow-up Actions.

---

[32] See الهيكل والتوصيات المتعلقة بإعداد تقرير الجودة لإحصاءات دول مجلس التعاون لدول الخليج العربية GCC-Stat, 2018 https://www.gccstat.org/images/gccstat/docman/Standards/tawsyat.pdf.

Additional reports may also be needed. For example, the main operational phases (Fieldwork, Processing, Macro Editing and Dissemination) will have regular reports on progress. These should include a summary of the quality issues and the relevant actions.

Quality reports may also be required for specific phases. These include Fieldwork Errors, Feedback from Public via Social Media, Call centre reports, as well as error reports from Processing, Macro Editing or Dissemination.

Administrative Censuses will require additional reports, for example on the quality issues of registers. The checklists described in Section 3 will be a major source for these reports.

4.2 Audiences

Overview

Proper and regular reporting helps provide a common understanding between stakeholders (including those providing administrative register data), decision makers, NSO managers and technical staff. The census project typically is conducted over a number of years, so reporting is important to help maintain the visibility of the project.

Within the census project, regular reporting helps with coordination, especially for those working on different phases and activities across the census cycle. This can also be useful for keeping the project team motivated.

However, not audiences need the same information or in the same format. While it is important for Census reporting to link into standard NSO reporting systems, some stakeholders may need specific reports. The stakeholders for the Census include Users, Suppliers (including agencies who are providing administrative records), Management and Staff within the NSO, and the Project governance group such as the Higher Level Committee.

Users

Any census project has a range of users. These can include:

- **Specialist Users** are often sophisticated users, requiring detailed data and associated metadata. They may be Academics, Researchers or come from international and regional organisations. These users will often use a range of statistics and so will need detailed quality information (e.g. error rates) to help decide if specific census outputs are appropriate for their needs.
- **Professional Users** include people from Government agencies, Libraries, and Businesses. They will make use of statistics on a semi-regular basis. These users need to understand the main quality issues to determine which statistics are suitable for their requirements.
- **General users**, such as the Public and Students, who make limited use of statistics. This may be via a third party, e.g. through social or traditional media. These users may have limited requirements for quality information.

Some users may sometimes be Professional Users – using some statistics intensely, but on other times being General Users – i.e. making limited use of statistical information. Senior Government Officials are examples of these types of users.

Quality reporting targeted to Users must therefore recognise the different types of users.

Suppliers

Suppliers, especially agencies who are providing administrative records may be interested in feedback on the quality of their register sources, as well as on the overall quality of the census results. These suppliers may have very targeted interests, focused on the issues of relevance to the agency.

NSO

The NSO is also made of many different stakeholders. Senior Management will generally be focused on the overall progress, the cost and the overall quality of the census statistics. Other people, such as Subject Matter Statisticians, Methodologists, IT team and Metadata team may be focused on their areas of expertise and involvement. Subject Matter Statisticians and Methodologists may be deeply interested in the quality of the census, as they determine how to maximise the value of the census data across the NSO.

High Level committee and Decision Makers

This forum is usually concerned with high-level issues, including the major quality issues that will affect the overall project, and relevant decisions. However, they may also require feedback on whether other audiences are being kept informed.

While all stakeholders have interests in the GCC Quality Reports, there are also audience specific requirements. These requirements also change across the census cycle. As noted above, some Users may be mainly interested in the overall quality of Outputs, and so are only engaged during the Dissemination phase. Other stakeholders such as Data Suppliers are mainly interested in the quality of their registers. The interest will be stronger during the Preparation and Collection/Collation phases, but there will also be some interest during the Dissemination phase.

Reporting therefore need to reflect the specific interests of the different types of stakeholder, which will vary in strength across the census phases. Figure 18 summarises the main interests for the different stakeholder groups (audiences) across the census cycle.

Figure 18: Quality Reporting interests for Census Stakeholders across the Census Cycle

	Census Phase					
	Preparation	**Collection/ Collation**	**Processing**	**Analysis**	**Dissemination**	**Evaluation**
Audiences						
Users	Processes				Processes	GCC Quality Reports
					Administrative Sources	
					Final Outputs	
Suppliers	Processes				Processes	GCC Quality Reports
	Administrative Sources				Administrative Sources	
					Final Outputs	
NSO	Processes					GCC Quality Reports
	Administrative Sources				Administrative Sources	
					Final Outputs	
High Level Committee	Processes					GCC Quality Reports
		Administrative Sources			Administrative Sources	
					Final Outputs	

Keys

Strength of Interest			Topics of Interest
Very Strong			Processes
Strong			Administrative Sources*
Moderate			Final Outputs
Some			GCC Quality Reports

*Only applies to Administrative or Combined Census

Appendices

Appendix 1: Administrative Census Quality Checklist

Countries in the GCC are conducting an Administrative Register based Census in the 2020 round. Administrative data requires detailed assessments to assess the quality.

These checklists have been prepared to help NSOs in the GCC to evaluate the quality of the administrative registers sources. These checklists follow the Hyperdimension approach – separating out checks on sources, metadata and the final data items

Hyperdimension	Definition	Purpose of Checks
Source	The information about the register as a whole, the owner and the environment for its creation, updating, access and use.	Understand the environment for the creation and ongoing operation of the register.
Metadata	The information about items in the register, including definitions and classifications.	Identifying what metadata is available to the NSO, understanding this information and determining whether it meets the needs of the NSO.
Data	The observed facts in the data	Understand in detail the data items and units. This will help determine the sources for the data items to be included in the final Census dataset. The assessment will also help refine the methods used to prepare this dataset.

Using the Hyperdimension approach means that checks are first made about the overall register (Source), to determine if the Register is suitable for use in the Census. Subject to the results, the Metadata is separately reviewed to check if the Register is suitable to be used in the Census. The Source and Metadata checks also provide important information to assist in the checks on the Data.

Together these checks help the NSO:
- determine what registers and data items to use,
- prepare and refine the methods (statistical, methodological and operational) needed to prepare the Census statistical dataset,
- Prepare the statistical metadata (conceptual, process and quality) needed to help users and the NSO understand the final Census results.

This approach allows for a progressive understanding of the registers and sources. Restrictions on the use of particular registers (e.g. legal or access constraints), can be identified before any detailed analysis is made of the data. Early identification and understanding of the documentation and metadata facilitates more effective investigation and analysis of the data.

Checklists
There are five checklists:

- **Source Checklist** – checks of each potential administrative register source. These checks are <u>conducted with the Register Owner,</u> with some information obtained within the NSO
- **Metadata Checklist** – checks of metadata for each suitable source. These checks are <u>conducted with the Register Owner</u>
- **Data Checklist** – detailed checks on the data, including population units, identifiers, data items and their respective values. Includes specific checks for <u>base registers and specialist sources.</u> Some checks <u>may require input from the Register Owner</u>.
- **Linking Checklist** - reviews whether the linking methods have worked as planned. These checks <u>are only conducted by the NSO</u>.
- **Statistical Dataset Checklist** – final check by the NSO that the final administrative based dataset meets the census requirements.

The checklists include Recommended Actions, Review and Decision Points to help ensure a quality census dataset.

Instructions

Each checklist covers several quality areas. Each area contains a series of indicators, scored by filling one or more questions. In some cases, the response is descriptive (e.g. describing the register owner's processes). In other cases, the response is numerical.

While most information is obtained from the Register Owner (the agency responsible for the register), some information also needs to be obtained directly from the NSO. A number of checks also involve comparisons of information provided by the Register Owner with the requirements of the NSO.

Most checks in the Data Checklist include an indicator. These can help determine the relative quality of data items from different sources.

Specific Notes

Actions are shown in red, as in the following example.

2.5.3	***Overall comments on Statistical Metadata (2.3)*** Specifically consider:
	• Availability and clarity of definitions of Population Units and Data Items.
	➢ Discuss any Data Items with a rating of Description Missing or Description Unclear with the Register Owner to obtain the necessary information.
	➢ If information on the definitions of the Population Units and Data Items is not available or clear, the Register is **NOT SUITABLE** for use in the Census.

Missing or Don't Know responses - Any responses of "Don't Know" or Missing, must be clarified with the Register Owner or NSO, and completed, before finalising the checklists.

NSO specific input - areas where information is only needed from the NSO, are shown as NSO only

Routing instructions are shown as "GO TO".

1. Register Source Checklist

Information about the Register and Contact People

Element	Instruction
Register	
Register Name	Name of Source – including internet address, if applicable
Date of Assessment	
Register Owner	
Name of Organisation	Agency who has responsibility for the register
Address	Physical Location
Postal address	
Website/ social media addresses	
Agency Contact Person	
Name	
Role/Responsibility	
Function and organisational unit	Department of contact person
Telephone Number	
Email address	
NSO Contact Person	
Name	
Role/Responsibility	
Function and organisational unit	
Telephone Number	
Email address	
Expected Role in Census	
Likely Role of Register	What is the Expected Role of this Register in the Census? 1. Part of Population Base Register (Fully or in part) 2. Part of Address/Housing Unit Base Register (Fully or in part) 3. Specialist Register source 4. Combination – specify
Details of Assessor (s)	
Name of Assessor(s)	
Title	
Date of Assessment	

	Check	Required Information/Indicator
1.1	**Reason for Administrative Register**	
1.1.1	What is the scope/purpose of the Source?	Describe key purpose and role of Register
1.1.2	What administrative processes are used to create and update the Administrative Register?	Describe the main processes, including who provides the original information, how the register is updated
1.1.3	Is the register used to report on performance targets?	Describe the role of the register in regard to performance measurements and targets.
1.2	**Legal and Security**	
1.2.1	What is the legal basis for the Administrative Register?	Include reference to the law, decree or legal agreement
1.2.2	Does the NSO have a legal mandate to obtain register data from the agency? (NSO to answer)	1. No 2. Yes
1.2.3	Does the agency have a legal obligation to provide data to the NSO?	1. No 2. Yes - Include reference to relevant section of law, decree or legal agreement 0. Don't Know
1.2.4	Does the agency have any restrictions on availability/access/ use of data/metadata?	1. No 2. Yes – Specify 0. Don't Know
1.2.5	Are there any specific data security requirements?	1. No 2. Yes – Specify 0. Don't Know
1.2.6	Does the NSO need to purchase any special hardware and/or software to enable secure delivery?	1. No 2. Yes – Specify 0. Don't Know
1.3	**Current Experiences (NSO only to complete)**	
1.3.1	Does the NSO use the <u>individual records</u> from the Administrative Register?	1. No – do not use this Register – GO TO 1.4 2. Only use aggregate records – GO TO 1.3.3 3. Yes, use individual records – Specify 0. Don't Know GO TO 1.4
1.3.2	Does the NSO combine the individual records with other data (other administrative data or survey data)?	1. No – GO TO 1.3.3 2. Yes – Specify 0. Don't Know
1.3.3	Are the terms of delivery documented ?	1. No 2. Yes – single general contract or MOU 3. Yes – specific contract or MOU 0. Don't Know
1.3.4	How often is the data delivered?	1. On request 2. Regular intervals
1.3.5	How punctual is the current delivery	1. Delivery varies – specify 2. Always on time GO TO 1.3.8

1.3.6	Are delays reported?	1. No – NSO not informed 2. Yes – NSO informed in timely manner GO TO 1.3.8 3. Yes – NSO informed, but not in timely manner
1.3.7	What arrangements are made when the data source is not or only partially delivered on time?	Describe arrangments
1.3.8	Is the NSO allowed to ask questions or contact the register owner in case of problems?	1. No – specify reasons 2. Yes – specify any constraints on feedback
1.3.9	How effective is the contact and communication with the agency?	1. Not effective 2. Very effective
1.3.10	What quality (eg accuracy, timeliness) issues have been found?	
1.4	**Content/Coverage of Administrative Register**	
1.4.1	What data can be delivered to the NSO?	List population units (groups covered by the register) and data items (variables)
1.4.2	Is the register known to be missing people or addresses?	1. No – GO TO 1.4.6 2. Yes – Specify 0. Don't Know
1.4.3	If yes, does the Data Owner plan to fill the gaps?	1. No 2. Yes – Specify 0. Don't Know
1.4.4	Does the register have duplicate records?	1. No 2. Yes – Specify 0. Don't Know
1.4.5	How many records (approximately) does the register currently contain?	
1.4.6	Does the agency plan to make changes to the register, including content (data items, classifications, etc), coverage, processes and/or data sources?	Describe any plans, including expected dates for changes.
1.5	**Identifiers**	
1.5.1	Does the source have identifiers (unique keys that can be used to identify the population units)	1. No 2. Yes - specify 0. Don't Know
1.5.2	Are there any restrictions on the NSO use of the Identifiers?	1. No 2. Yes - specify 0. Don't Know
1.5.3	Does the Administrative Register use standard government Identitiers? (Eg ID card numbers for people)	1. No - Specify the agency identifiers 2. Yes – Specify which standards. GO TO 1.5.5 0. Don't Know GO TO 1.5.5

1.5.4	Can the agency specific identifiers be mapped to the relevant government standard ?	1. No 2. Yes 0. Don't Know
1.5.5	What data management practices are used for identifiers?	Identify all practices (eg use check digits, special checks for duplicate identifiers)
1.5.6	Are there combinations of data items that can be used to uniquely identify population units ?	1. No 2. Yes – specify the combinations 0. Don't Know 9. Not applicable
1.6	**Use for Statistical Purposes**	
1.6.1	Is the Source currently used by the Register Owner to produce statistics?	1. No 2. Yes – specify 0. Don't Know
1.6.2	Are the data items in the register defined?	1. No 2. Yes – specify 0. Don't Know
1.6.3	What classifications are used in the register?	Specify all classifications– eg ISIC Rev 4
1.6.4	Have definitions or classifications changed over time?	1. No 2. Yes – specify 0. Don't Know
1.6.5	Does each record have a geography?	1. No 2. Yes - specify level and geographic classification 0. Don't Know 9. Not applicable
1.6.6	Is the Source used to provide sub-national (governate,etc) information?	1. No 2. Yes – describe process 0. Don't Know
1.7	**Timeliness and Punctuality**	
1.7.1	Will information be able to be extracted for the census reference date?	1. No GO TO 1.10 2. Yes 0. Don't Know
1.7.2	How frequently is the register updated?	
1.7.3	When was the last update?	
1.7.4	How long does it take for the register to be updated after the relevant event?	
1.7.5	Are there any plans to change the timeliness?	1. No 2. Yes 0. Don't Know
1.7.6	How quickly after the census reference date, would the updated data would be available ?	

1.8	**Accessibility and Transfer**	
1.8.1	Is the register currently available electronically?	1. No 2. Yes GO TO 1.8.3 0. Don't Know GO TO 1.8.3
1.8.2	When will the register be available electronically?	
1.8.3	What is the planned Data transfer process?	Describe the proposed manner (eg Government Network, e-mail, DVD, etc)
1.8.4	What format will the data be provided in?	1. Agency specific format (Specify) 2. Standard Government formats 0. Don't Know
1.8.5	What is the proposed frequency of transfer for the census project?	
1.8.6	Are there any direct costs to use the Administrative Register?	1. No 2. Yes – specify 0. Don't Know
1.9	**Relationships with Agency (NSO only to complete)**	
1.9.1	What relationship does the NSO have with the Register Owner?	Describe the relationship (Eg regular senior level meetings)
1.9.2	Is there a MOU with the agency?	1. No 2. Yes – specify 0. Don't Know
1.9.3	What influence does/can NSO have on the Administrative Register? (eg content, timeliness?)	
1.9.4	Does the Census High Level Committee/Steering Committee includes the agency?	1. No 2. Yes – specify 0. Don't Know
1.10	**Summary (NSO only to complete)**	
1.10.1	**Overall comments on Reason for Administrative Register (1.1)** Specifically consider • If the processes used to create/maintain the register will provide statistical data (1.1.2) • If the register is used to measure performance (1.1.3), could that impact on the use for statistical purposes	
1.10.2	**Overall comments on Legal and Security (1.2)** Specifically consider • Possible legal issues (1.2.1 - 1.2.3) • If there are any possible contradictions between the NSO legal mandate and agency laws. (1.2.2 and 1.2.3) • Agency security or privacy requirements (1.2.4 -1.2.6)	
1.10.3	**Overall comments on Current Experiences (1.3)** If there are any issues with the Source, specifically consider: • Any possible impacts on the census	

1.10.4	**Overall comments on Content/Coverage of Administrative Register (1.4)** Specifically consider: • Coverage of register (units and data items) (1.4.1) • Coverage gaps (1.4.2 and 1.4.4) • Does the estimate of number of records on the register(1.4.5) match other information. If not, does this match the owner's views on coverage (1.4.2) • Agency plans to change coverage (1.4.6) ➤ What is the likely impact on the coverage of the Census? If there are significant coverage gaps, the source **may not** be suitable.
1.10.5	**Overall comments on Identifiers (1.5)** Specifically consider: • Availability of identifiers (1.5.1) • Restrictions on the NSO use of Identifiers (1.5.2) • Ability to match Agency identifiers to the Government standard (where relevant) (1.5.4) • Availability of combinations of data items that will uniquely identify each record? (1.5.6) ➤ Note the **Source is NOT SUITABLE in** the following circumstances: • no unique identifiers or unique combinations of data items, • restrictions on the use of identifiers mean it is not possible to uniquely identify each record, including assigning statistical identifiers
1.10.6	**Overall comments on Use for Statistical Purposes (1.6)** Specifically consider: • Lessons from current statistical uses (1.6.1) • Comparability of definitions and classifications (1.6.2, 1.6.3) • Geographic coding (1.6.5, 1.6.6)

1.10.7	**Overall comments on Timeliness and Punctuality (1.7)**
	Specifically consider:
	  • Whether information is available for the Census Reference Date (1.7.1)
	➤ If information is not available for the Census Reference date, or can be converted to match with the Reference Date; then the Register Source is NOT SUITABLE
	  • Timeliness of availability/reporting (1.7.1 – 1.7.4, 1.7.6)

1.10.8	**Overall comments on Accessibility and Transfer (1.8)**
	Specifically consider
	  • Availability of electronic records (1.8.1, 1.8.2)
	  • Transfer methods (1.8.3- 1.8.5)
	  • Current Transfer and availability issues (1.3.3 -1.3.7)

1.10.9	**Overall comments on Relationships with Agency (1.9)**
	Consider
	  • If the processes in place with the Agency are appropriate for issues identified in these Source checks
	  • What changes, if any, are needed?

Decision Point 1 Suitability following Source Checks

Review/ Decision	Use the Checklist 1.10.1 to 1.10.9, to determine whether the Register source is suitable for the census
	    1. **No**
	    2. **Partly – specify**
	    3. **Yes, as planned**

Action	➢ **If the checks show that the Source is suitable (Fully or partly), proceed to the Metadata checks** ➢ **Otherwise, do not conduct any further investigation of the source.** ➢ **In all cases, document the results of the Checklist.**

2. Metadata Checklist

These include checks on the information about the register, as well as checks on the operations (processes) conducted by the Register Owner. Information may be recorded in writing, or provided verbally. (Where information is provided verbally, it should be carefully documented.)

	Check	Required Information/Indicator
2.1	**Availability and Clarity**	
2.1.1	How will the NSO understand the contents of the register? If no information will be available (written or verbal), the register is NOT SUITABLE and no more checks are needed.	1. No Documentation or Verbal briefings are available – GO TO 2.5 2. Only Verbal briefings will be provided - GO TO 2.1.4 3. Documentation fully or partly available 0. Don't Know – GO TO 2.5
2.1.2	Is the documentation clearly structured and well organised	1. No 2. Yes 0. Don't Know
2.1.3	Are there any plans to change the documentation?	1. No 2. Yes - specify plans 0. Don't Know
2.2	**Agency Processes**	
2.2.1	Does the information show how the register is created and updated? This may include data entry/receipt of data from other agencies.	1. No 2. Yes 0. Don't Know
2.2.2	Does the Register have a formal process to track changes? (e.g. time and change owner)	1. No - specify how changes are recorded 2. Yes 0. Don't Know
2.2.3	Does the Register Owner check the population units?	1. No 2. Yes – specify types of checks 0. Don't Know
2.2.4	Does the Register Owner check data items (e.g. range checks)?	1. No 2. Yes – specify types of checks 0. Don't Know
2.2.5	Does the Register Owner check the plausibility of combinations (e.g. validation checks)?	1. No 2. Yes – specify types of checks 0. Don't Know
2.2.6	Does the Register Owner check for the occurrence of extreme values	1. No 2. Yes – specify types of checks 0. Don't Know
2.2.7	Does the Register Owner modify (edit, impute) data?	1. No GO TO 2.2.9 2. Yes – specify types of actions 0. Don't Know
2.2.8	Are modified values marked in the Register?	1. No 2. Yes – original data included or available 3. Yes – but original data not available 0. Don't Know

2.2.9	Is there information about the treatment of special cases?	1. No 2. Yes – specify what is available 0. Don't Know
2.2.10	Does the Register Owner have any other data management and validation processes?	1. No 2. Yes – specify what is available
2.2.11	Does the Register Owner carry out internal/external audits on the data or on the processes used to produce the data in the register?	1. No 2. Yes – Audits conducted on data and processes (Specify) 0. Don't Know
2.3	**Statistical Metadata**	
2.3.1	Are all the population units defined clearly?	1. Description unclear 2. Yes – description clear 0. Description missing GO TO 2.3.3
2.3.2	Describe the population units as defined by the Register Owner	
2.3.3	Are all the data items defined?	1. No - Missing for some - specify gaps 2. Yes 0. Definitions missing for all / Don't Know, GO TO 2.3.6
2.3.4	Are data item definitions clear? (Complete the assessment for **each data item planned to be used in the census.**)	1. No - Desciption unclear/ambiguous 2. Yes – description clear 0. Description missing
2.3.5	Are definitions of all time periods included? (Complete the assessment for **each data item assessed in 2.3.4***)	1. Description unclear/ambiguous 2. Yes - description clear 0. Description Missing 9. Not applicable
2.3.6	Are classifications included in the information about the register? (Complete the assessment for **each data item assessed in 2.3.4**)	1. No 2. Yes 0. Description Missing 9. Not applicable
2.3.7	Are code files for any agency specific coding available? (Complete the assessment for **each data item assessed in 2.3.4**)	1. No 2. Yes 0. Description Missing 9. Not applicable
2.3.8	Are changes in definitions, classifications or code files, recorded?	1. No - specify gaps 2. Yes 0. Don't Know 9. Not applicable
2.4	**Consistency and Comparability (NSO to complete)**	
2.4.1	How comparable are the definitions of the population units used by the register owner and the NSO?	0. Description missing/ not available 1. Definitions unequal – conversion is impossible 2. Definitions unequal – conversion is possible 3. Equal (100% identical)

2.4.2	How comparable are the definitions of the data items? (Complete for **each data item** assessed in **2.3.4**.)	0. Description missing/ not available 1. Definitions unequal – conversion is impossible 2. Definitions unequal – conversion is possible 3. Equal (100% identical)
2.4.3	Are the time periods, including reference periods comparable? Complete **for each applicable data item assessed in 2.3.5**	0. Description missing/ not available 1. Definitions unequal – conversion is impossible 2. Definitions unequal – conversion is possible 3. Equal (100% identical)
2.4.4	Are the classifications comparable? Complete **for each applicable data item assessed in 2.3.6**	0. Description missing/ not available 1. Classifications not consistent – unable to create concordance file 2. Classifications not consistent – able to create concordance file 3. Equal (Same classification, including version)
2.5	**Summary (NSO to complete)**	
2.5.1	**Overall comments on Availability and Clarity of Metadata (2.1)** Specifically consider: • Availability of Metadata and the options to understand the contents of the register (2.1.1) ➢ No available information (written or verbal) means that the Source **is NOT SUITABLE**	
2.5.2	**Overall comments on Agency Process Documentation (2.2*)* Specifically consider: • Availability of process information (2.2.1 – 2.2.3) and the options to understand the processes, if there is limited available metadata.	
2.5.3	**Overall comments on Statistical Metadata (2.3***)* Specifically consider: • Availability and clarity of definitions of Population Units and Data Items. ➢ Discuss any Data Items with a rating of Description Missing or Description Unclear with the Register Owner to obtain the necessary information. ➢ If information on the definitions of the Population Units and Data Items is not available or clear, the Register is **NOT SUITABLE** for use in the Census.	

| 2.5.4 | **Overall comments on Consistency and Comparability (2.4).** Specifically consider

• Population Units (2.4.1), Data Items (2.4.2), Time Periods (2.4.3) and Classifications (2.4.4) which are rated as 'description missing' or 'unequal and conversion is impossible"

➤ Discuss any Data Items with a rating of Description Missing or "Unequal and conversion is impossible" with the Register Owner. If information is not available, or it is not possible to convert them to the statistical standards, then these Data Items are **NOT SUITABLE**

• Consider whether the definitions of **Population Units (2.4.1), Data Items (2.4.2)** and Classifications (2.4.4) are consistent with census requirements.

➤ If the definitions of **Population Units and/or Data Items** are not consistent with census requirements, determine if there are technical options to transform the data. Where this is not possible, the **source** is **NOT SUITABLE** for use in the Census

➤ If the **classifications** are not comparable, and it is not possible to prepare concordances or recode the original data, then the **data item will NOT BE SUITABLE** for use in the census. Other data items may be suitable. |

Decision Point 2 Suitability following Metadata Checks

Decision / Review	Use the Checklist to determine whether the Register source is suitable for the census
	
	   1. No
	   2. Partly – specify which population units and data items are suitable
	   3. Yes

Action	➤ **If the checks identify that the Source is suitable (Fully or in part), proceed to the Data checks.** ➤ **Otherwise, do not conduct any further investigation of the source.** ➤ **In all cases, document the results of the Checklist.**

3. Data Checklist

These are the detailed checks on the data, including population units, identifiers, data items and their respective values. These data checks require some data analysis, including comparisons with other data sources. The checklist separates out:

- Completeness checks – whether the provided data is consistent with the Source checklist.
- Consistency checks – identifies any major inconsistencies between the data and the Metadata checklist.
- Detailed Investigations of Data – assesses individual data item level, or combinations of data items.

Completeness checks

	Check	Required Information/Indicator
3.1	**Accessibility and Transfer**	
3.1.1	Was the register extract provided according to the agreed file transfer arrangements?	• Proportion of register extract consistent with file transfer arrangements
3.1.2	Can records be read electronically?	• Proportion of Units Records that can be read electronically
3.1.3	Was the size of the transferred file consistent with expectations?	• Ratio of size of file received to size of file expected
3.2	**Availability and Clarity**	
3.2.1	Is the register extract consistent with agreed requirements?	• Number of Missing Population Units • Ratio of number of Data Items supplied to number of Data Items expected • Number of missing data items • Overall consistency (Yes, No)
3.2.2	Does the register extract include records not required or requested by the NSO? (eg records related to a different time period or different set of population units.) Does the register extract include data items not required or requested by the NSO?	• Number of extra records supplied • Number of extra data items supplied

Data Review Point 1 - Completeness

Review	Confirm if the Register extracts are consistent with Source Checklist
Action	➤ **If the extract does not meet requirements, contact the Register Owner.** ➤ **Repeat Completeness checks (3.1 -3.2), if a new extract is provided.**

Consistency Checks

	Checks	Required Information/Indicator
3.3	**Detailed Population Unit checks (for all Population Units planned to be used in the Census)**	
3.3.1	Is the population unit consistent with the definition in the metadata?	• Proportion of records with inconsistent population unit definitions
3.3.2	If the metadata specified any special cases, are they included in the extract?	• Number of special cases in metadata missing from the register extract
3.3.3	Are any special cases found in the extract, identified in the metadata?	• Proportion of special cases not described in the metadata
3.4	**Detailed Data Item checks (separate checks for <u>each Data Item</u> assessed as Suitable in Decision Point 2)**	
3.4.1	Is the data item consistent with the definitions in the metadata information?	• Proportion of records with inconsistent data item definitions
3.4.2	Is each data item consistent with the validation rules provided in the metadata? (SEE also Checks 2.2.4 – 2.2.9)	• Proportion of records with data items inconsistent with the validation rules in metadata • Number of data items with at least one record failing the validation rules in the metadata 0. Validation rules missing in metadata 9. Not applicable - No validation undertaken
3.4.3	For each coded data item, are the code values consistent with the code files/classifications provided in the metadata?	• For each coded data item, proportion of records with missing codes • For each coded data item, proportion of records with invalid codes • For each coded data item, proportion of records with codes out of range 9. Not applicable – Data Item not coded
3.5	**Time Period Checks**	
3.5.1	Are the time periods in the data extract, consistent with requirements? (SEE also Check 2.4.3) Are the time periods consistent with the metadata?	• Proportion of records with time periods inconsistent with requirements • Proportion of records with time periods inconsistent with the metadata
3.5.2	Do the data items relating to events have relevant event dates?. (Eg Marriages should have a Date of Marriage)	• Proportion of data items relating to events which do not have relevant event dates.
3.5.3	Are reference dates for events (where relevant), identified?	• Ratio of records missing required reference dates to records requiring reference dates 9. Not applicable – GO TO 3.5.6
3.5.4	Are the reference dates for data items consistent with the metadata?	• Proportion of records with inconsistent reference dates

3.5.5	Do any of the events have event dates after the reference date (E.g. Births after the Reference Date)	• Proportion of records which have event dates after the reference date.
3.5.6	Does the register contain time periods without events (eg months without marriages or births)? Are these consistent with the Metadata?	• Number of unexplained Time Periods without events

Data Review Point 2 – Consistency with Metadata

Review	Review whether the provided data is consistent overall with the metadata
Action	➢ **If the extracts are not consistent with the Metadata, contact the Register Owner.** • **Specify which data items and Population Units are consistent with the Source and Metadata Checklists**

Detailed Investigation of Data

Core Data Checks

	Check	Required Information /Indicator
3.6	**Identifiers**	
3.6.1	Do all records have an identifier?	• Proportion of records without any identifier
3.6.2	Do all records have valid identifiers?	• Proportion of records with invalid identifier
3.6.3	Is the identifier consistent with the definitions in the metadata? (SEE Check 1.5)	• Proportion of records with identifier definitions inconsistent with the metadata
3.6.4	Are there any records with duplicate identifiers, but different data items?	• Proportion of records with duplicate Identifiers, but different data item values
3.6.5	Are any records fully duplicated in the register(ie same identifier and same data item values)?	• Proportion of records that are fully duplicated
3.6.6	For registers with incomplete identifiers, does the register contain unique combinations of variables to identify units?	• Unique combination of variables to identify units exists (Yes, No)
3.6.7	If required, can Statistical Identifiers be created from Administrative (Government or Agency) Identifiers?	• Proportion of records not possible to create Statistical Identifiers from Administrative Identifiers
3.6.8	Where unique combinations of variables exist, rather than Identifiers, is it possible to create Statistical Identifiers?	• Proportion of records not possible to create Statistical Identifiers from Unique combinations of records
3.7	**Coverage**	
3.7.1	Are there records with identifiers, but no data item values?	• Proportion of records with identifiers, but no data item values
3.7.2	Are there records with data item values, but no identifiers	• Proportion of records with data item values, but no identifiers
3.7.3	Are there population units missing from the register? Is the register extract biased? (E.g. regional, ages, nationality, etc.)	• Estimate of number of population units that appear to be missing from the register • Estimate of bias in the missing population units
3.8	**Content (Required for each data item confirmed in Data Review Point 2)**	
3.8.1	Does each record have values for each data item? Are the missing values randomly distributed, or is there a bias?	• For each data item, proportion of records missing relevant values • Estimate of bias in the missing values Requires comparison with other sources.
3.8.2	Do any data items have invalid record formats? (eg alpha response, when number is expected, Age when Date of Birth expected, Month and Year when full Date of Birth expected, etc)	• For each data item, proportion of records with invalid formats

3.8.3	For each data item, are all the values within range?	• For each data item, proportion of records with range errors
3.8.4	Are the records internally consistent?	• Proportion of records with internal consistency errors
3.8.5	For each coded data item, are the codes consistent with the code files/classifications provided in the metadata?	• For each coded data item, proportion of missing codes • For each coded data item, proportion of invalid codes
3.8.6	For all data items recorded as text, what proportion do NOT have codes assigned?	• For each relevant data item, Number of records required to be coded
3.8.7	Are the data items consistent with other statistical information for the same variables? (Eg are the distributions consistent?)	• For each data item, consistency of statistical information (Yes, No)

Data Review Point 3 – Quality of Data

Review	• **Review the quality indicators, specifically** o **Confirm if Indicators exist or can be created** o **Review if the coverage of each <u>Population Unit</u>, meets census needs** o **Whether the quality indicators for each data item meets census needs**
Action	➢ **Discuss systematic quality issues with the Register Owner. Examples include records with missing or invalid identifiers, high levels of missing data** ➢ **Update the list of suitable Population Units and Data Items, recorded in Data Review Point 2.**

Specific Checks on Base Registers

Base Registers (Address Base Register, Population Base Register) form the base of the Administrative Register Census and should have 100 % coverage.

	Check	Required Information/ Indicator
3.9	**Sources for Address Base Register**	
3.9.1	Does each housing unit (including institutions, worker camps and temporary housing units) have an address? Are the records without an address randomly distributed, or is there a bias? (E.g. Are the missing records in specific location or specific housing unit type)	• Proportion of records without an address. • Estimate of bias in missing address values • Are areas of bias consistent with the metadata?
3.9.2	Is the geography code used in the register source, consistent with the metadata?	• Proportion of records with geographic codes inconsistent with the metadata

3.9.3	Does each housing unit have a geography code? Are the missing values randomly distributed, or is there a bias? (E.g. location, housing unit type)	• Proportion of records without a geography code. • Estimate of bias in missing geography code values • Are areas of bias consistent with the metadata?
3.9.4	Does each housing unit with a geographic code also have an address?	• Proportion of records with geographic code, missing addresses.
3.10	**Final Address Base Register**	
3.10.1	Can the Address Base Register be created?	• Address Base Register Creation (Yes, No)
3.10.2	Does the Address Base Register include: • Addresses for all housing units • Geographic codes for all addresses	• Estimate of coverage of Address Base Register • Estimate of coverage of geographic coding of Address Base Register
3.11	**Sources for Population Base Register**	
3.11.1	Are population units (people) missing from the register source? Is the register extract biased? (E.g. regional, ages, nationality, etc.)	• Estimate of number of population units (people) that appear to be missing from the register • Estimate of bias in the missing population units • Are areas of bias consistent with the metadata?
3.11.2	Are any gaps in the coverage of the Register able to be filled by another source (e.g. another Register/ Fieldwork)?	• Sources identified to fill gaps in the register? (Yes, No) 9. Not applicable, register covers 100% of population
3.11.3	Can the base register source be linked to the existing base register?	• Source able to be linked to existing base register (Yes, No) 9. Not applicable, only one source used for Base Register
3.12	**Final Population Base Register**	
3.12.1	Can the Population Base Register be created?	• Population Base Register Creation (Yes or No)
3.12.2	Does the Population Base Register include records for all Usual Residents?	• Estimate of number of population units (people) that appear to be missing from the register

Data Review Point 4 – Base Registers

Review	• **Review the Base Register quality indicators, to assess if quality Base Registers can be created**
Action	➢ **If either the Population or Address Base Registers cannot be created, review how the Administrative Census will work.(E.g.it may be necessary to move to a Combined Census – and use the Fieldwork to prepare the missing Base Register.)** ➢ **If the Base Registers can be created, but there are quality issues, identify additional sources.** ➢ **In all cases, document the decisions and assessments in the relevant metadata.**

Specific Checks on Extracts from Specialist Registers

Specialist registers are sources of specialist information about people or housing units used to provide information about their characteristics.

	Checks	Required Information/ Indicator
3.13	**Extracts from Specialist Registers (Assess for each specialist register source.)**	
3.13.1	Identify whether the records can be linked with the relevant Base Register.	• Proportion of records that cannot be linked with the relevant Base Register

3.13.2	Are the specific data items on the Specialist registers appropriate for the census? (Consider results from the Content checks (3.8))	• Accuracy and Consistency indicators such as
		   o Proportion of missing values
		   o Error rates
		   o Coding errors

Data Review Point 5 – Specialist Registers

Review	• **Review the quality indicators for each Specialist source. Specifically consider**
	   o **Indicator and Coverage checks**
	   o **Whether records can be linked to the base registers..**

Action	➢ **Determine Specialist source to determine which data items can be used.** ➢ **When all of the Specialist sources have been assessed, update the list of suitable Population Units and Data Items** ➢ **In all cases, document the decisions and assessments in the relevant metadata.**

Decision Point 3 Suitability following Detailed Data Checks

Decision / Review	Use the Checklists to determine whether it is technically feasible to conduct an Administrative based census 1. No 2. Partly – specify which population units and data items will need to be obtained, or validated through Fieldwork 3. Yes
Action	➢ If the checks identify that it is technically feasible (partly or fully), to conduct an Administrative census, complete the Linking checks. ➢ If it is not feasible, it will be necessary to find an alternative method to conduct the census. ➢ In all cases, document the results of the Checklist.

4. Linking Checklist

These checks review whether the linking methods have worked as planned. The checklist also identifies the overall success of the linking process.

	Check	Required Information/ Indicator
4.1	**Linking Methods**	
4.1.1	What linking methods were used to create the Address Base Registers?	• Linked source registers using o standard government identifiers o statistical identifiers o Combination of government and statistical identifiers o Other Linking methods - specify • Converted Administrative register to Statistical Address Base Register
4.1.2	What linking methods were used to create the Population Base Registers?	• Linked source registers using o standard government identifiers o statistical identifiers o Combination of government and statistical identifiers o Other Linking methods - specify • Converted Administrative register to Statistical Population Base Register
4.1.3	What methods were used to link Specialist register sources to the Address Base Register?	• Linked registers using o standard government identifiers o statistical identifiers o Combination of government and statistical identifiers o Other Linking methods - specify 9. Not applicable as no Specialist sources were needed
4.1.4	What methods were used to link Specialist register sources to the Population Base Register?	• Linked registers using o standard government identifiers o statistical identifiers o Combination of government and statistical identifiers o Other Linking methods - specify 9. Not applicable as no Specialist sources were needed
4.2	**Linking of Population and Housing Registers**	
4.2.1	Can the records on the Population Register be linked to the Address Base Register?	• Proportion of Population Base Register records that can not be linked to a Housing Unit/Address • Estimate of bias in the records that cannot be linked
4.2.2	Can the records on the Address Base Register be linked to the Population Register?	• Proportion of occupied Housing Units that cannot be linked to the Population Register • Estimate of bias in the records that cannot be linked

Decision Point 4 – Review Linking

Review	**Review the quality indicators for Linking. Do the indicators confirm that the linking worked as planned?** **Are the results of the review of Linking consistent with the investigations in 3.10, 3.12 and 3.13?**
Action	**If the linking has not worked as expected, review the linking methodology and data sources** **Once the linking is working as required, conduct the final Statistical Dataset Creation checks.** **In all cases, document the decisions and assessments in the relevant metadata.**

5. Statistical Dataset Creation

These check the final dataset against the requirements.

	Check	Required Information/ Indicator
5.1	**Coverage**	
5.1.1	Does the Population Base Register include all the population groups required by the census? (I.e. does it cover the Usual Resident Population?)	• Proportion of the Census Usual Resident Population covered by the Population Base Register • Estimate of under/over coverage • Estimate of bias in the missing population units (people)
5.1.2	Does the Address Base Register include all the housing units, required by the census, including institutions, worker camps and any temporary units?	• Proportion of required Census Housing Units covered by the Address Base Register • Estimate of under/over coverage • Estimate of bias in the missing housing units (eg region, occupancy status, tenure)
5.2	**Integration**	
5.2.1	Does the Statistical Dataset contain all the necessary people, household/family and housing unit information?	• Proportion of People on the Statistical Dataset who do not have a geographic code • Proportion of occupied Housing Units, who do not have usual occupants • Proportion of occupied Housing Units where it is not possible to determine the Housing type (Family/Share/Collective) • Proportion of family and share housing units, where it is not possible to identify the household members
5.2.2	Can all the required data basket topics (including derived topics), be produced?	• Number of Population topics in the Census Data basket that cannot be produced • Number of Housing topics in Census Data Basket that cannot be produced
5.2.3	Does the level of substitute records meet requirements?	• Number of substitute records created for Addresses/Housing Units • Number of substitute records created for Individuals/People
5.2.4	Has Reference Metadata been created for the integrated Census dataset?	• Availability of Reference Metadata (Yes/No)
5.3	**Content**	
5.3.1	Are the values of all data items in the required range?	• Proportion of records with values not in the required range
5.3.2	What errors are in the coded data items?	• Estimated level of recoding errors for pre-coded data items • Estimated number of text fields requiring coding or recoding

5.3.3	Are the records within the dataset consistent?	• For each data item, proportion of records which needed to be changed to ensure consistency
5.3.4	What is the level of Item Non-response?	• Level of item non-response for each data item • Proportion of data items meeting minimum Non-response standards

Decision Point 5 – Review of Final Integration Dataset

Review	• **Review the final quality indicators for the Integrated Dataset and confirm that it meets requirements.** • **Has Reference Metadata been created for the final dataset?**
Action	➢ **If the Integrated dataset is of acceptable quality, then it can now be used in the next step of the census cycle.** ➢ **If the quality is not acceptable, determine whether the underlying issues are the quality of the administrative sources or the methodologies used to derive the census.** ➢ **If the full set of Reference Metadata has not been created, actions must be taken to complete the documentation – including completing the checklists** ➢ **In all cases, document the decisions and assessments in the relevant metadata.**

General Terms

Base Registers	The Statistical registers relating to People (Population Base Register) and Housing/Addresses (Address Base Register). Another example of a Base Register is the Establishment Register. This is the register of all establishments in the country.
Classification	A set of related categories used to group the data according to its similarities. Examples include ISIC, ISCED.
Codes	Data items may be recorded using text, numbers or alphanumeric codes. The numbers or alphanumeric codes may relate to a classification.
Data Item	The items included in the register.
Event	A record may relate to an event, e.g. arrival in the country, birth, death, marriage, renting a property.
Event date	Events will also have an event date, for example date of birth, start of rental contract.
Government Identifiers	Identifiers used by Government agencies to uniquely identify individuals, household, families, housing units, addresses or similar units. Government identifiers include standard Government identifiers used across many government administrative systems, (e.g. Government ID number) and Agency specific identifiers. Agency specific identifiers relate to one agency's administrative system. Student ID numbers used in the Ministry of Education are an example of agency identifier.
Housing Units	Places of Accommodation. These include private housing units used by families, other accommodation such as hotels and camps and worker's quarters
Identifiers	Identifiers uniquely identify each record. The identifiers may be standard Government identifier (e.g. Government ID number), Agency specific identifiers (e.g. Student ID number) or statistical identifier.
Population units	The specific group covered by the register. Examples include all people, people over 15, citizens only, families only, all buildings, rented housing units.
Record	Entry in a register relating to an individual, household, family or housing unit. It will usually contain some form of identifier and values for a number of data items (variables)
Register	A register is a systematic collection of unit-level records, each containing an identifier or identifiers and relevant data items (variables), organized so that updating is possible.
Register Extract	This refers to the records from the administrative register provided for use in the Census.
Register Owner	The organisation responsible for the register. The government normally operates administrative registers. However, administrative registers may also be operated by private organizations.
Registration date	The date when the event is recorded in the register
Statistical Data Set	The Census dataset, prepared by linking Base Registers with the different Register Extracts from Specialist Administrative Registers.

Statistical Identifier	The identifier used in the statistical system to uniquely identify records. The identifier may be created from the Government or Agency Identifier, or by matching a unique combination of variables (e.g. name +date of birth + mother family name + ...)
Specialist Registers	Administrative Registers used to provide specific information about people or housing units.
Updating of Register	The processing of identifiable information with the purpose of establishing, updating, correcting or extending the register.
Usual Residents	All the people (citizens and non-citizens) who usually live in a country, that is, who have lived continuously in a country for most of the last 12 months (i.e. for at least six months and one day), not including temporary absences for holidays or work assignments; or intend to live in that country for at least the next six months

Metadata Terms

Documentation	Existing information about the register provided by the register owner.
Metadata	Metadata describe statistical data and the processes and tools involved in the production and usage of statistical data. Metadata for administrative censuses is created from different places, including: • documentation and information provided by register owners about the individual registers, • information obtained by the NSO in assessing the register sources and their data • documentation of actions and processes undertaken by the NSO
Reference metadata	Metadata describing the contents and the quality of the statistical data. This is prepared to users understand and use the final statistics. Ideally, it should include all of the following: a) "conceptual" metadata, describing the concepts used and their practical implementation, allowing users to understand what the statistics are measuring and, thus, their fitness for use; b) "methodological" metadata, describing methods used for the generation of the data (e.g. sampling, collection methods, editing processes); c) "quality" metadata, describing the different quality dimensions of the resulting statistics (e.g. timeliness, accuracy).

Data Manipulation Terms

Consistency errors	Errors can occur between data items, when the value of one data item is not consistent with others.
Derived variables	A new data item created from other data elements using a mathematical, logical, or other type of transformation, e.g. arithmetic formula, composition, aggregation. An example for individuals is labour force status, derived from responses to data items for employment, unemployment. Derived variables can also be aggregates such as unemployment rates, total population.

Errors	Errors are invalid values or inconsistencies in records. Errors can be at the data item level or between data items. Errors can result in the system stopping or the resulting information to be inconsistent or invalid
Expected Error Rates	The expected threshold/benchmark set by the NSO for the proportion of errors or changes to records from repairing errors.
Fatal errors	These errors identify major inconsistencies with the data. They must be fixed. Examples include records with invalid geography codes.
Linking	Linking aims to join records relating to the same population unit (persons, housing units/addresses), but coming from different sources, using a common identifier or combination of data items. Data Linking can be either **deterministic** or **probabilistic**. **Deterministic** linking uses either unique identifiers or an exact combination of data items. **Probabilistic** matching is used where no single data item can provide a reliable match. Several data items are compared between two records and each data item is assigned a weight that indicates how closely the two values match. The sum of the individual weights indicates the likelihood of a match.
Non-response errors	These include cases where the data item has a missing value.
Query edits	These are cases where it is suspected that the value of the data item or combination of data items is invalid. An example is a person where the education and occupation do not match.
Range errors	These are where the data item has a value, but it is outside the specified range.
Structural errors	Those errors that cause the programme/system to stop, and so must be fixed before proceeding with this data. An example is an address with a status of occupied, but no person records.

Appendix 2: GCC Quality Assessment Templates

GCC-Stat recommends that the dissemination of GCC statistics be accompanied by a quality description. The quality description is a concise assessment of the quality in term of the reliability and relevance of the statistics for different purposes and user needs. Its main objective is to show what approach and methods are applied, and how the quality criteria are fulfilled. This ensures transparency in quality evaluation and quality assurance. It is recommended that quality reports be published at the same time as the concerned statistics. This allows users to be informed about limits and constraints of statistical information and process.

Recommended Structure:

The quality report for the 2020 Census is based on the GCC standard as set out in الهيكل والتوصيات المتعلقة بإعداد تقرير الجودة لإحصاءات دول مجلس التعاون لدول الخليج العربية (Recommendations for preparing quality assessment reports of GCC statistics). It should include the following:

1. Introduction

An introduction for the context of quality reporting, including:

- A brief history of the Census, its methodology and approach, and main outputs
- Reference to other documentation (questionnaire, methodology…)

2.   Information on each of the quality components

a. Relevance of statistical information

Relevance is the degree to which statistical outputs meet current and potential user needs at national, GCC, regional and international levels*.*

b. Accuracy of data

Data accuracy refers to the degree of closeness of estimates to the true values.

c. Timeliness and punctuality of data

The timeliness of statistical outputs is the length of time between the event or phenomenon they describe and their availability. The punctuality is the time lag between the release date of data and the target date on which they were scheduled for release.

d. Accessibility and clarity of data

Accessibility and clarity refer to the simplicity and ease with which users can access statistics, with the appropriate supporting information and assistance such as metadata, documentation, explanation, quality limitations, etc...

e. Comparability and coherence of data

Coherence refers to the degree to which data derived from different sources but measuring the same phenomena are similar to the estimates generated by the program.

Comparability refers to the degree to which statistical outputs refer to the same data items and the aim can be comparable over time, or across regions, or across other domains.

3. A short assessment of User Needs and Perceptions

Description of the main uses, and users, and their feedback

4. Conclusion

Section 1 Identification

Name of NSO	
Division/ Department responsible	
Name of Collection – i.e. 2020 Population and Housing Census	
Contact Person	
Name	(Person 1), (Person 2)
Role	(Person 1), (Person 2)
Telephone Number	(Person 1), (Person 2)
Email address	(Person 1), (Person 2)
Census Methodology	
Reference Date	
Census Collection Methodology	1. Administrative only (No Fieldwork) 2. Combined – Administrative and Fieldwork 3. Fieldwork only
Population Base	1. Usual Residence Census 2. Census Night Count 3. Mixture of Usual Residence and Census Night

Section II: Quality Reporting

1 Relevance
This section should provide the following:
• Summary of the information content and purpose/use of the Census statistics
• Introduction to the concepts and definitions and associated classifications
• Description of output products/services at different levels of detail, formats and media

Relevant Quality Indicators

Indicator	Description	Questions
1.1 Rate of available statistics	Meeting international and GCC Stat data requirements	1. Number of data items missing from the GCC 2020 Census Data Basket 2. List of missing data items 3. Number of data items that do not meet the GCC Data Basket concepts and definitions 4. List of data items with different concepts and/or definitions 5. Number of missing outputs (e.g. tables) from the agreed GCC 2020 Census outputs 6. List of missing outputs
1.2. Products and Services	Availability of Products and Services	1. What Products and Services were provided from the 2020 Census? (Please list)
2 Accuracy		

This section should include the following:
- Processes used to assess accuracy, including Quality Control and Quality Assurance processes used in all steps. Include information on the Macro-Editing (Analysis) and Evaluation processes

- Possible sources of error – including identified errors from Fieldwork, Processing, or Dissemination. Information from Macro-editing and Evaluation may be also be used.
- Results of the evaluation of the census, including information on coverage
- Documentation of any errors identified once the Census statistics had been published.
- Description of any revisions made to the Census statistics, including differences between Initial (Preliminary) and Final release of statistics
- Description of the Disclosure Control (confidentiality protection) used on outputs.
- Description of other confidentiality protection practices used in the Census process

Relevant Quality Indicators (Note 2.2 Applies only to Administrative based Censuses)		
Indicator	**Description**	**Questions**
2.1 Testing	Tests conducted	1. List the tests conducted in the 2020 Census

2.2 Administrative Census	Specific accuracy measures of Administrative Census	1. Number of Registers which passed the Assessments
		  1. Sources checks
		  2. Metadata checks
		  3. Data checks
		2. Number of data items which passed Assessment checks
		  1. Metadata checks
		  2. Data checks
		3. Proportion of required units covered by the Population and Address Base Registers
		4. Main linking methods used to :
		  1. Create Address Base Register
		  2. Create Population Base Register
		  3. Link Specialist sources to the Population Base Register
		  4. Link Specialist sources to the Address Base Register
		  5. Link Address and Population registers
		5. How necessary was it to edit the Register records?
		  1. A lot of errors were discovered and checking and editing was indispensable
		  2. Few errors needed to be corrected
		  3. Records had already been sufficiently checked and were error free. No data editing was necessary

2.3. Over-coverage in Final Statistics	Units (People and Addresses) included in the final Census dataset that do not belong.	1. What is the level of over-coverage? (Report separately for People and Addresses)
		  1. Not measured
		  2. Major over-coverage
		  3. Some over-coverage
		  4. Slight over-coverage
		  5. Other (please specify)

2.4. Under-coverage in Final Statistics.	People and Addresses that are not in the final Census dataset, but should have been	1. What is the level of under-coverage? (Report separately for People and Addresses)
		  1. Not measured
		  2. Major under-coverage
		  3. Some under-coverage
		  4. Slight under-coverage)
		  5. Other (please specify)

2.5. Edit failure	Records that triggered error	1. How necessary was it to edit the records?
		  1. No data editing was necessary
		  2. A lot of errors were discovered and checking and editing was indispensable
		  3. Few errors needed to be corrected
		2. What methods were used to identify errors
		  1. Not required
		  2. Automated editing
		  3. Computer Assisted editing
		  4. Manual editing
		  5. Combination
		  6. Other – please specify

2.6 Coding	Records that required coding	1. How necessary was it to code the records?
		  1. No coding was necessary
		  2. A large number of data items needed to be coded
		  3. Some data items needed to be coded
		  4. Only one or two data items needed to be corrected
		2. What methods were used to identify errors
		  1. Not required
		  2. Automated coding
		  3. Computer Assisted coding
		  4. Manual coding
		  5. Combination
		  6. Other – please specify

2.7 Unit response rate.	Units in the Final database which are missing some, but not all data items	1. What was the level of unit non-response? (Address/ Housing Units and People)
		  1. Unit non–response rate is not known or unacceptably high
		  2. High non-response rate (more than 15%)
		  3. Medium non-response rate (5 % up to 15 %)
		  4. Low non-response rate (less than 5%)
		  5. There is nearly no unit non-response

2.8. Item response rate.	Key data items, which did not have responses from all units. (Key variables are location, age, sex, nationality.)	1. What was the level of item non-response for key variables? 1. Rate of item non–response is not known 2. Rate of item non-response is unacceptably high (>50%) 3. There is a lot of item non-response (15% to under 50%) 4. There is some item non-response (5% to under 15%) 5. There is little item non-response (<5%) 6. There is nearly no item non-response
2.9. Imputation	Assigning values where data is missing, invalid or inconsistent	1- How necessary was it to impute values? 1- No imputation was necessary 2- A large number of records had to be imputed 3- Some records required values to be imputed 4- Only a few records needed to be imputed 2- What methods were used for imputation 1- Not required 2- Automated system 3- Computer Assisted system 4- Manual system 5- Other – please specify
2.10. Substitute records	Substitute records are created when an address or person is known to exist, but no records have been received	1. What level of substitute records were created? (Address/ Housing Units and People) 1. Not measured 2. A large number of records had to be created 3. Some records had to be created 4. A small number records had to be created 5. No substitute records had to be created 6. Other (please specify)
2.11. Macro Editing	Analysis of aggregates and patterns in the outputs	1- Describe the major issues identified in Macro-Editing

| 2.12 Errors in Dissemination | Mistakes identified during the production of outputs or after their release | 1- Were errors identified during the preparation of outputs?
  1- Not measured
  2- A large number of errors were found
  3- Some errors were found
  4- A small number of errors were found
  5- No errors were found
  6- Other (please specify)
2- Were errors identified in published census statistics?
  1- Not measured
  2- Large number of errors were reported
  3- Some errors were reported
  4- Small number of errors were reported
  5- No errors were reported
  6- Other (please specify) |

3 Timeliness and Punctuality

This section should include the following:
- The release schedule for all the Census output products and services, including which deadlines were achieved
- The web address(es) where users can find the release calendar and census outputs
- Feedback from users on the timeliness of the Census statistics

Relevant Quality Indicators

Indicator	Description	Questions

3.1. Time lag for release of first/provisional results	The number of days from the reference period to the day of publication of first results.	1- What is the time lag between the census reference date and the first publication of the preliminary results?
		  1- Substantial time lag
		  2- There is a certain time lag (within the agreed time schedule)
		  3- There is a small time lag
		  4- Other (please specify)

3.2. Time lag for release of final results	The number of days from the reference period to the day of publication of final results.	1- What is the time lag between the reference period and the first publication of the final results?
		  2- Substantial time lag
		  3- Some time lag, but within the agreed time schedule)
		  4- There is a small time lag
		  5- Other (please specify)

3.3. Punctuality of publication.	Actual release date meets the previously announced date of publication	1- Was the Release calendar for census usually kept?
		  1- No release calendar
		  2- There was frequently a large delay
		  3- There was sometimes a certain delay
		  4- Dates in calendar always kept
		  5- Other (please specify)

4 Accessibility and Clarity		
This section should include the following: • Different types of products and services provided from the census. • Where users can find the description, metadata, and quality and methodological descriptions of the statistics • Feedback from users on accessibility and clarity		
Relevant Quality Indicators		
Indicator	**Description**	**Questions**

4.1. Subscriptions/ purchases of each of the key reports	Number of subscriptions/purchases of each of the key paper reports	1. Do you regularly keep track of subscriptions/purchases of each of the key paper reports?
		  1. Don't produce paper reports
		  2. Yes, track subscriptions/ purchases of reports
		  3. Produce paper reports, but do not track subscriptions/ purchases

4.2. On-line products and services	Access and use of on-line products and services	1. Do you regularly keep track of the usability of your website for different groups of users?
		  1- Don't release census data on-line
		  2- Provide census data on-line, but don't track usability
		  3- Yes, track usability of census data
		2. Number of accesses to on-line databases or on-line products
		3. Do you invite user comments on the content and presentation of your statistical outputs?
		4. What proportion of census 2020 data is available on-line?
		  1. Census Data is not available on-line
		  2. A large part of the data is on-line
		  3. A certain part of data is on-line
		  4. A small part of data is on-line

4.3 Metadata	Availability and Completeness of Metadata	1. What metadata is provided to users?
		  1. Metadata is not provided
		  2. Information on methodology is provided
		  3. Information on quality of statistical processes and products is provided
		  4. Information on methodology and quality of statistical processes provided using GCC-Stat metadata template
		2. What metadata is available for data items?
		  1. Metadata is not provided
		  2. Information provided for all data items
		  3. Information provided for most data items
		  4. Information provided for a few data items
		  5. Other, specify

5 Coherence and Comparability		
This section should include the following:		
• Uniformity and coherence of the statistics relative to other statistics on the same topic. This will be based on the results of Macro-Editing (Analysis) and Evaluation		
• Comparisons of the 2020 Census with previous censuses or other time series. This will also be based on the results of Macro Editing and Evaluation.		
Relevant Quality Indicators		
Indicator	**Description**	**Questions**
5.1. Comparability	Comparability of census with other statistics and over time.	1. How comparability were the census statistics with other statistics? 　1. Comparability not assessed 　2. Serous differences 　3. Large differences 　4. Some differences 　5. Slight differences 　6. No major differences 　7. Other (please specify) 2. How comparable are the statistics from the 2020 Census with previous censuses? 　1. Not assessed 　2. Not comparable over time due to fundamental changes 　3. Not comparable over time due to problems 　4. Comparability over time is seriously restricted 　5. Comparability over time is restricted to some data items 　6. Comparability over time is slightly restricted 　7. Fully comparable
5.2. Coherence of key estimates with other estimates	Comparison of the key estimates with estimates from alternative sources	1- Did you conduct any Evaluation studies on the Census? 　1. No Evaluation studies were done 　2. Conducted Post Enumeration Survey 　3. Completed Demographic Analysis 　4. Compared Census with other statistics 　5. Other – specify 2. How coherent are the Census statistics with other sources, including important non-official statistics? 　1. Not assessed 　2. Substantial differences 　3. Lot of differences 　4. Some differences 　5. No major differences

5.3 Bias	Impact of errors on Final statistics	1- What is the overall impact of the errors on the 2020 Census statistics? 1- Not Assessed 2- Little or no impact 3- Some impact – some statistics have bias 4- Major impact – many statistics have bias 5- Severe impact – overall census has bias

6 Assessment of User Needs and Perceptions

This section should include the following:

- Description of how the user perspective was considered in the preparation of the Census outputs, including user consultation
- Summary of the main uses of Census statistics, including level (e.g. geography, other main breakdowns). Also include information on the use of the different census products and services
- Feedback from users on availability of the Census statistics
- Identification of areas where it was not possible to meet user requirements in the 2020 Census round

Relevant Quality Indicators

Indicator	Description	Questions
6.1. User satisfaction index	The degree of satisfaction with services and products for different segments of users.	1. What information is available on user satisfaction of the 2020 outputs? 1. Not assessed 2. Complete (detailed) information is available to users 3. A lot of information is available to users 4. Some information is available to users 5. Little information is available to users 2. Do you establish meetings/contacts with key users to study their needs of statistics (including new and priority needs)? 1- Regular meetings/contacts are carried out 2- Casual meetings/contacts are carried out 3- Meetings/contacts are not carried out **4-** Other (please specify)

7 Summary

This section should include the following:

- Main Quality Problems encountered in the 2020 Census
- Impact of quality problems on statistics and implications for the users
- Recommendations for improvement
- Follow-up actions

Appendix 3: Resources

Daas, P. et al, (2009) Checklist-quality-evaluation-administrative-data-sources, http://ec.europa.eu/eurostat/documents/64157/4374310/45-Checklist-quality-evaluation-administrative-data-sources-2009.pdf/24ffb3dd-5509-4f7e-9683-4477be82ee60

Eurostat, (2011) European Statistical Code of Practice for the National and Community Statistical Authorities, http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF/5fa1ebc6-90bb-43fa-888f-dde032471e15

GCC-Stat (2015) Pre-conditions for Administrative Register Census, http://gccstat.org/ar/elibrary/publications/gccstat/item/gcc-pre-conditions-for-an-administrative-census-in-gcc-countries-2

GCC-Stat (2017) Population Statistics in GCC Countries 2010- 2014, https://gccstat.org/en/statistic/publications/population-statistics-in-gcc-countries-2010%E2%80%932014

GCC-Stat (2018) Guideline to Planning and Preparing Administrative Register based censuses in GCC Countries

GCC-Stat (2018) الهيكل والتوصيات المتعلقة بإعداد تقرير الجودة لإحصاءات دول مجلس التعاون لدول الخليج العربية (Recommendations for preparing quality assessment report of GCC statistics) https://www.gccstat.org/images/gccstat/docman/Standards/tawsyat.pdf

GCC-Stat, (2018) إطار ضمان جودة البيانات الإحصائية لدول مجلس التعاون لدول الخليج العربية (Data Quality Framework for GCC Statistics) https://www.gccstat.org/images/gccstat/docman/Standards/b00k.pdf

Guerin, L. (2014) Editing the 2011 Census data with CANCEIS and options considered for 2016. Working paper, UNECE Work Session on Statistical Data Editing, https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.44/2014/mtg1/Topic_4_Canada_Guertin.pdf

IMF (2012), Data Quality Assessment Framework, https://dsbb.imf.org/dqrs/DQAF

ISO (2015) ISO 9001, https://www.iso.org/iso-9001-quality-management.html

Office of National Statistics (2012) Item Edit and Imputation: Evaluation report, http://webarchive.nationalarchives.gov.uk/20160107185728/http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/index.html

Office of National Statistics, (2016) Evaluating the potential for moving away from a traditional census, Conference of European Statisticians Group of Experts on Population and Housing Censuses Eighteenth Meeting Geneva, 28 - 30 September 2016, https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2016/mtg1/CES_GE.41_2016_18E.pdf

Scottish Government Social Research Group (2009), Cognitive Testing in Survey Questionnaire Design - Social Science Methods Series Guide 7: http://www.gov.scot/Resource/Doc/175356/0091403.pdf

Software Testing Fundamentals (undated) Software Development Life Cycle, http://softwaretestingfundamentals.com/software-development-life-cycle

Statistics Canada (2009) Statistics Canada Quality Guidelines, Fifth Edition - Collection, Capture and Editing, http://www.statcan.gc.ca/pub/12-539-x/2009001/collection-collecte-eng.htm

Statistics Canada (2009) Statistics Canada Quality Guidelines- Data Dissemination and Communication, http://www.statcan.gc.ca/pub/12-539-x/2009001/dissemination-diffusion-eng.htm

Statistics New Zealand (2014) Understanding substitution and imputation in the 2013 Census, http://archive.stats.govt.nz/Census/2013-census/methodology/substitution-and-imputation.aspx

UK Statistics Authority (2015) Quality Assurance of Administrative Data – Setting the Standard, https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-settingthestandar_tcm97-44370.pdf

UK Statistics Authority, (2015) Administrative Data Quality Assurance Toolkit, version 1 https://www.statisticsauthority.gov.uk/wp-content/uploads/2015/12/images-qualityassurancetoolki_tcm97-44368.pdf

United Nations Commission for Europe (2007) Register-based Statistics in the Nordic Countries –Review of best Practices with focus on population statistics, http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

United Nations Commission for Europe, (2013), Generic Statistical Business Processing Model *Version 5, December 2013*, http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5

United Nations Statistical Division, (2014) Fundamental Principles of National Official Statistics, https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx

United Nations Statistics Division (2001) Handbook on Census Management for Population and Housing Censuses, No 83, Revision 1 https://unstats.un.org/unsd/censuskb20/Attachments/Census%20Management-e-GUIDfd30f57d4e0d43f8b4101f1cb3266b7a.pdf

United Nations Statistics Division (2010) Handbook on Population and Housing Census Editing Revision 1, https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1e.pdf,

United Nations Statistics Division (2017) Principles and Recommendations for Population and Housing Censuses – Revision 3 https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles_and_Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf

Wallgren, A & Wallgren B (2014) *Register-based Statistics: statistical methods for administrative data*, Second Edition, 2014, United Kingdom Wiley Series in Survey Methodology