

EGM on the Revision of the P&R

Session 10

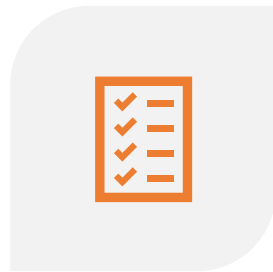
TT5: Data collection and processing

15 December 2023

Our Starting Point...

- TT5 was to review and update the parts of the P&R that elaborate on:

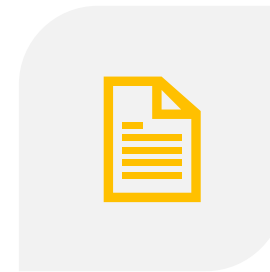
TT5: Data collection and processing



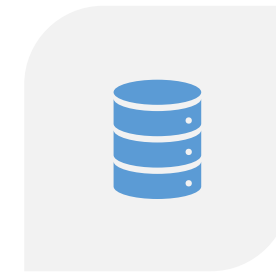
CENSUS QUESTIONNAIRE
DESIGN



BUILDING AND TESTING
CENSUS INFRASTRUCTURE



FIELD ENUMERATION



DATA PROCESSING



Scope of Work

The scope of work of TT5 will include (but not limited to):

- review and update chapters II, III, V, VI, VII, VIII of part III of the P&R pertaining to census questionnaire design, building and testing census infrastructure, field enumeration and data processing
- provide guidance on designing questionnaires for electronic data collection (CAPI, CAWI, CATI)
- address the importance of implementing real-time editing (both soft and hard edits) at the time of data collection when using electronic data collection technologies
- provide guidance for the implementation of short and long questionnaires, potential operational challenges and recommendations on how to deal with challenges, such as integration of data collected from different questionnaires and ensuring consistency in data processing
- provide guidance for the design and implementation of a multi-mode data collection, and for ensuring coherence among different data collection modes, and for dealing with mode effects
- discuss the effects of electronic data collection on data processing and data quality, and provide guidance for merging data from different collection modes
- discuss the role of geospatial information in monitoring and managing field enumeration
- discuss the role of alternative data sources (such as administrative data and big data) during data collection (e.g., preloading information into electronic questionnaire) and during data processing (e.g., for validation and imputation purposes)

Into subgroups...

- August formation into subgroups
- 1 September 2023 →
 - Subgroup 1 will revise topics related to **“Census questionnaires: content and design”**
 - Subgroup 2 will revise topics related to **“Building census infrastructure / Census tests / Living quarters and household listing”**
 - Subgroup 3 will revise topics related to **“Field enumeration”**
 - Subgroup 4 will revise topics related to **“Data processing”**

Team consisted of...

15+ individuals

4 subgroups

| Subgroup 1 | Subgroup 2 | Subgroup 3 | Subgroup 4 |
|---|---|------------------------------------|-----------------------------|
| Census questionnaires: content and design | Building census infrastructure / Census tests / Living quarters and household listing | Field enumeration | Data processing |
| Lead: Margaret Mbogoni (UNSD) | Lead: Jennifer Reichert (United States) | Lead: Sergiy Radyakin (World Bank) | Lead: Paul Waweru (Kenya) |
| Seif Kuchengo (Tanzania) | Sergiy Radyakin (World Bank) | Paul Waweru (Kenya) | Luqmaan Omar (South Africa) |
| Papa Ibrahima Sene (Senegal) | Papa Ibrahima Sene (Senegal) | Ebtesam Al Shehhi (UAE) | Fariza Hanim (Malaysia) |
| Fredrick Okwayo (UNFPA-ESA) | Ebtesam Al Shehhi (UAE) | Jennifer Reichert (United States) | Lenin Aguinaga (UNECLAC) |
| Oliver Fischer (IPC/USCB) | Oliver Fischer (IPC/USCB) | Seif Kuchengo (Tanzania) | |
| Sandile Simelane (UNFPA-HQ) | Seiffe Tadesse (UNSD) | Fredrick Okwayo (UNFPA-ESA) | |
| Seiffe Tadesse (UNSD) | | Sandile Simelane (UNFPA-HQ) | |
| Fariza Hanim (Malaysia) | | Margaret Mbogoni (UNSD) | |
| Lenin Aguinaga (UNECLAC) | | Luqmaan Omar (South Africa) | |

Our key objective

01

Revision of existing text

02

Recent lessons learned, innovations and good practices included

03

Is the guide adequate for the 2030 round censuses?

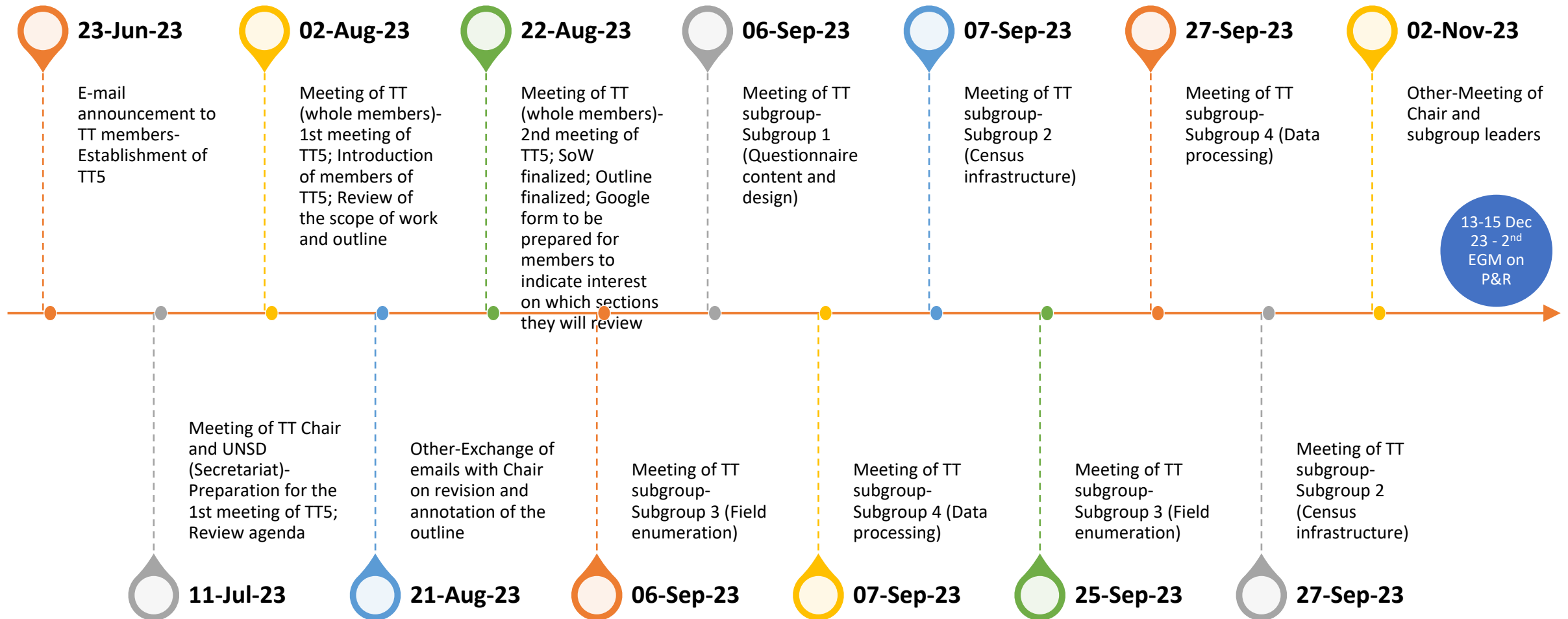
Summary of Chap 3 – When we started

153
paragraphs

10 sections

| | | | Start | End | Total |
|-------|--|----|-------|-----|-------|
| I. | Introduction | 2 | 1 | 1 | 1 |
| II. | Census questionnaires: content and design | 2 | 2 | 35 | 34 |
| III. | Building census infrastructure | 7 | 36 | 43 | 8 |
| V. | Census tests | 8 | 44 | 48 | 5 |
| VI. | Living quarters and household listing | 10 | 49 | 52 | 4 |
| VII. | Field enumeration | 11 | 53 | 53 | 1 |
| A. | Method of enumeration | 11 | 54 | 68 | 15 |
| B. | Timing and length of the enumeration period | 14 | 69 | 74 | 6 |
| C. | Management and supervision | 15 | 75 | 87 | 13 |
| D. | Use of technology | 17 | 88 | 96 | 9 |
| VIII. | Data processing | 19 | 97 | 102 | 6 |
| A. | Method of processing | 20 | 103 | 105 | 3 |
| B. | Preparation for data capture | 21 | 106 | 107 | 2 |
| C. | Data capture | 22 | 108 | 115 | 8 |
| D. | Coding | 24 | 116 | 121 | 6 |
| E. | Data editing | 25 | 122 | 125 | 4 |
| F. | Validation | 26 | 126 | 131 | 6 |
| G. | Processing control | 27 | 132 | 133 | 2 |
| H. | Master file | 28 | 134 | 136 | 3 |
| I. | Methods of tabulation | 28 | 137 | 140 | 4 |
| XII. | Archiving individual records | 29 | 141 | 141 | 1 |
| A. | Purpose of archiving individuals records | 29 | 142 | 145 | 4 |
| B. | Procedures for archiving | 30 | 146 | 151 | 6 |
| C. | Archiving individual records and microdata | 31 | 152 | 153 | 2 |

Engagement Timeline



Where are we now...

- Document has grown (1008 revisions)
- 50-60% larger

| Word Count | |
|---|---------|
| Statistics: | |
| Pages | 32 |
| Words | 17 397 |
| Characters (no spaces) | 96 690 |
| Characters (with spaces) | 113 941 |
| Paragraphs | 262 |
| Lines | 1 471 |
| <input checked="" type="checkbox"/> Include textboxes, footnotes and endnotes | |
| Close | |

| Word Count | |
|---|---------|
| Statistics: | |
| Pages | 1 |
| Words | 26 346 |
| Characters (no spaces) | 151 007 |
| Characters (with spaces) | 177 224 |
| Paragraphs | 420 |
| Lines | 2 380 |
| <input checked="" type="checkbox"/> Include textboxes, footnotes and endnotes | |
| Close | |

Sections/Revised or New

| Chapter/Section/Sub-section of the P&R | Status |
|--|---------|
| PART THREE. CENSUS OPERATION ACTIVITIES | |
| I. Introduction | Revised |
| II. Census questionnaires: content and design | Revised |
| a) Selection of census topics | Revised |
| b) Use of short and long questionnaires | Revised |
| c) Sections of the questionnaire | Revised |
| d) Questionnaire design | Revised |
| Design of questionnaires for electronic data collection (CAPI, CAWI, | New |
| III. Building census infrastructure | Revised |
| V. Census tests | Revised |
| VI. Establishing the census frame | Revised |

Sections/Revised or New

| Chapter/Section/Sub-section of the P&R | Status |
|--|-------------|
| VII. Field enumeration | Revised |
| A. Method of enumeration | Revised |
| B. Multi-mode data collection | New |
| a) Designing a census multi-mode data collection | New |
| C. Use of administrative data during enumeration | New |
| D. Timing and length of the enumeration period | New |
| E. Management and supervision | Revised |
| a) Management information system | Revised |
| b) Supervising the enumeration | Revised |
| i. Using historical data | Revised |
| ii. Setting goals and targets | Revised |
| iii. Preparing policies and procedures | Revised |
| F. Security during data collection | New |
| G. Use of technology | Revised |
| a) Electronic questionnaire | Revised |
| i. Electronic questionnaire: interviewer-administered | Revised |
| ii. Electronic questionnaire: self-enumeration method | Revised |
| b) Handheld or mobile devices | Revised |
| c) Geographic information system (GIS) and geospatial informatic | Revised/New |
| d) Dashboards | New |
| e) Contact centre | Revised |
| f) Situation room | New |
| g) Short messaging service (SMS) | Revised |

Sections/Revised or New

| Chapter/Section/Sub-section of the P&R | Status |
|---|---------|
| VIII. Data processing | Revised |
| A. Method of processing | Revised |
| B. Preparation for data capture | Revised |
| C. Data capture | Revised |
| Data capture during enumeration | New |
| Data capture after enumeration | New |
| D. Coding | Revised |
| E. Data editing | Revised |
| F. Validation | Revised |
| G. Processing control | Revised |
| H. Master file | Revised |
| I. Methods of tabulation | Revised |
| J. Use of administrative data during data processing (for traditional censuses) | New |
| K. Security during data processing | New |

Proposals under discussion/Further Revision



Mixed Mode vs Multi Mode vs
Multimodal vs Multi Channel vs
OmniChannel data collection

Parallel vs Sequential Application

The degree of Integration

Switching Modes/Channels – Continuity between Modes
and Waves



Big Data is the new Buzzword like
Electronic Data Collection and CAPI
was in the 2020 round

Big data can be a highly valuable complementary source
Recommendations need to be made guiding countries
that Big Data is an 'enabler' and a complementary source
but differing maturity states in countries and
"benchmarked" census methodologies need to be
followed to ensure the census principles are not violated



Thank You!

Members of TT5 Team

Special mention to Seiffe Tadesse
(Coordinator, Secretariat and Chief Doodle Poll Specialist)