

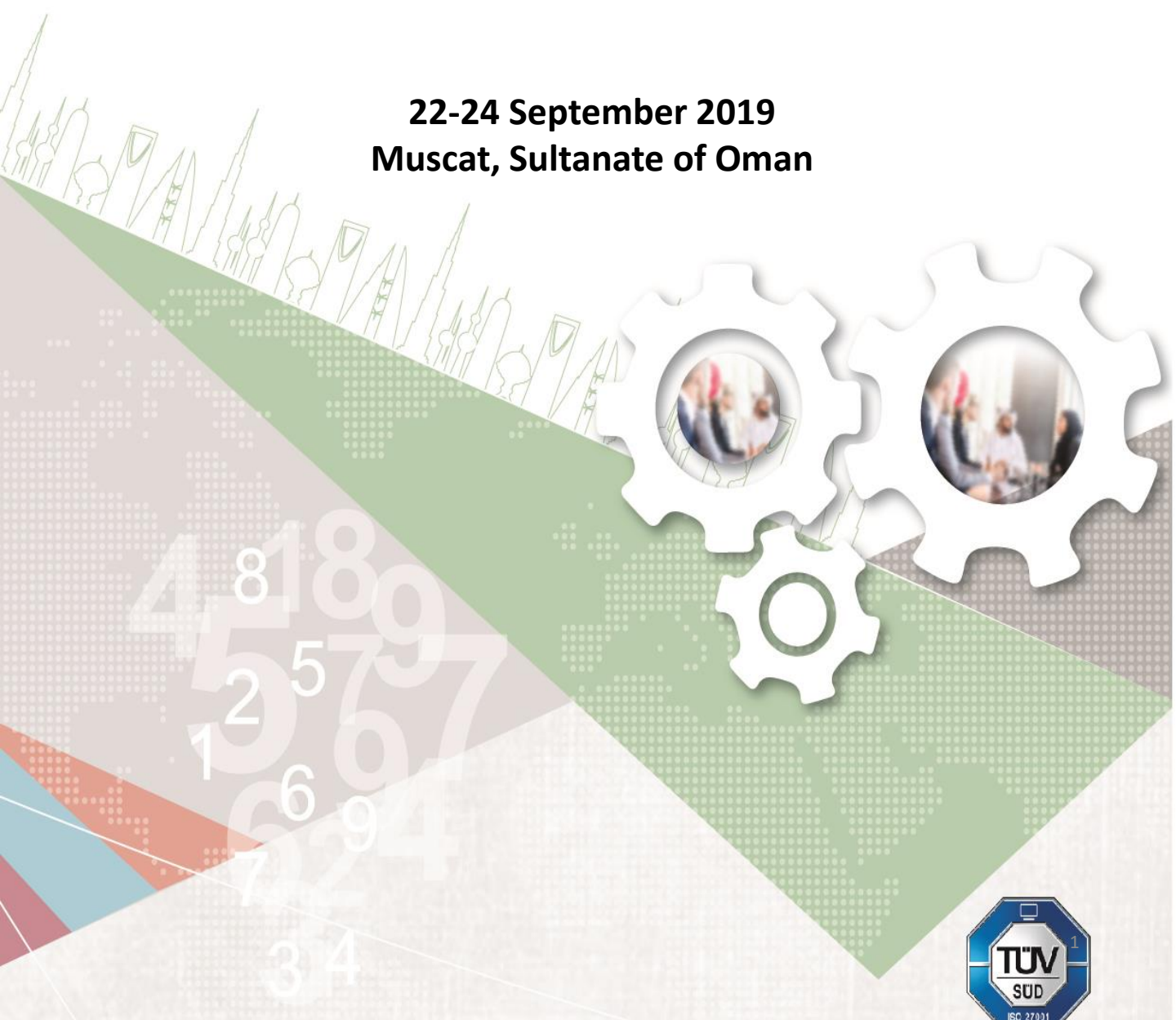


المركز الإحصائي
لدول مجلس التعاون لدول الخليج العربية
GCC-STAT



Regional Workshop on Editing (Data Cleansing) of Statistics produced from Administrative Registers

22-24 September 2019
Muscat, Sultanate of Oman





Preface

The GCC Regional Workshop is designed to share international best practices and regional experiences in the statistical editing (data cleansing) of unit record administrative data.

These techniques will be of particular value to NSOs in implementing the Strategic Plan for the Statistical Work of the GCC-STAT 2015-2020 for priority projects (2020 Census, Labor Statistics and Business Register).

1. Background

All countries in the GCC are making more use of administrative data in statistical work. All NSOs in the region will conduct an administrative based census in 2020. Many countries are now producing labour statistics from administrative records, as well as preparing administrative based business registers.

Administrative data is not however designed for statistical purposes. This means that NSOs have to transform the records to meet statistical needs. An important step in transforming administrative records is editing (data cleansing).

Editing has always had a key role in identifying and correcting errors and missing values that occur in the data collection process. In recent years, there has been considerable international research into statistical editing, in order to improve the efficiency and effectiveness of this critical process. Many of the international practices in editing of administrative data have drawn on this research, while also recognizing the specific issues associated with administrative data.

The workshop is being held to give practical support to NSOs in the use of unit record administrative data, especially in the 2020 Census, Labour Statistics and Business register projects.

The focus of the workshop is on editing/cleansing of unit record data. This workshop builds on the 2017 GCC Regional workshop on Census Data Quality, and the 2019 Workshop on Preparations for Census Data Evaluation, the GCC Regional workshop on labour statistics from administrative records in 2017 and country training on the use of Administrative data for Official Statistics (2017, 2019).



2. Main Themes

The workshop will address four key areas as follows:

Theme 1 Concepts of Editing of Administrative Data.

Theme 2 Quality of Administrative Data.

Theme 3 Editing (Data Cleansing) Techniques.

Theme 4 Implementing Editing Techniques.

3. Objectives and Aims of the workshop

The objective is to share knowledge in international best practice and techniques in the editing (data cleansing) of administrative based data. The specific aims of the workshop are:

- Understand international experiences in editing administrative data for statistical use.
- Review common types of errors found in administrative data.
- Understand role of statistical agencies and administrative agencies in the cleaning (editing) of administrative data for statistical purposes.
- Share experiences in common causes of errors in administrative data in GCC countries.
- Understand how to apply international best practice techniques in data editing including:
 - Correcting errors in record linking
 - Editing within individual register sources
 - Editing between different register sources
 - Cleaning of data that is not regularly updated by administrative agencies
 - Using Administrative data to help validate/correct survey data
 - Using survey data to help validate/correct administrative data
 - Review country experiences of editing (data cleansing), of administrative data



4. Outcomes

Participants will have a common understanding of principles and practices to be used in editing (cleansing) of administrative data.

5. Organization

The workshop has been organised by GCC-Stat, across three of the GCC statistical priority projects – Labour Statistics, Administrative Data and 2020 Population and Housing Census.

The workshop will be held at GCC-Stat, Muscat, Oman

6. Workshop Participants

The workshop is intended for statisticians and relevant IT staff working on the editing (cleansing) of administrative data in the GCC labour, 2020 census and GCC business register projects. The workshop will also be useful for senior officials in National Statistical Offices responsible for the management of these projects.

7. Speakers

- Mr. Danilo Dolenc, Head, Demography Statistics and Census Project Manager, Statistics Slovenia
- Mrs. Nancy McBeth, Consultant and Advisor, GCC-STAT
- Mrs. Elham Salah, Technical Support Advisor, GCC-STAT
- Ms. Batoul Obaid, Methodologies and Social Statistics Expert, GCC-STAT
- GCC Country representatives – tbc

8. Language

The workshop will use both Arabic and English, with simultaneous Interpretation



9. Requirements

Participants need to:

1. Review the concept paper.
2. Participants will need to present country experiences, and share experiences.
3. Participants are also requested to bring a sample of data, that they can work on during the practical sessions.

10. Contact Information

For further technical information, please contact

Mrs. Nancy McBeth – Consultant and Adviser – GCC STAT

Phone : +968 - 24346469

Mobile: +968 - 91706137

E-mail : nmcbeth@gccstat.org

Or

Mrs. Elham Salah- Technical Support Advisor – GCC-STAT

Phone : +968 - 24346441

Mobile: +973 - 36601113

E-mail : ESalah@gccstat.org

Or

Ms. Batoul Obaid – Expert – GCC-STAT

Phone : +968 - 24346461

Mobile: +968 - 97742411

E-mail : bobaid@gccstat.org

For further administrative and logistics, please contact

Mr. Thani Al Hadi, Head of the Protocol Section, GCC-Stat:

Phone: +968 - 24346432

Mobile: +968 - 99100011

E-mail: talhadi@gccstat.org



Workshop Agenda

22 September 2019

| Time | Speaker | Topic |
|-------------|-------------|--|
| 8:30 – 9.00 | | Registration |
| 9.00 – 9.30 | Elham Salah | Workshop Opening Opening Remarks Introductions and Expectations of Participants Workshop Objectives and Organisation |

Session 1: Concepts of Editing of Administrative Data

Using Administrative Data for Statistical Purposes in the GCC Statistical System

| | | |
|-------------|-------------|---|
| 9.30- 10.00 | Elham Salah | What is Administrative Data Overview of Administrative Data projects in GCC Statistical system |
|-------------|-------------|---|

Editing/Data Cleansing in the Data Cycle

| | | |
|--------------|--------------|--|
| 10.00 -10.45 | Batoul Obaid | Where editing fits in the Statistical Cycle Standard Type of Errors Coverage Error (including Missing/Invalid records) Content Error (Missing data values, Range errors, Inconsistencies) |
|--------------|--------------|--|

10.45- 11.00

Coffee Break

Structure of Administrative Data

| | | |
|---------------|---------------|--|
| 11.00 – 11.30 | Danilo Dolenc | -Identifiers and Linking Variables -Data Sources -Data Variables |
|---------------|---------------|--|

Roles of NSOs and Administrative Agencies

| | | |
|--------------|--------------|--|
| 11.30 -12.00 | Nancy McBeth | - Custodians, owners, users of administrative data |
|--------------|--------------|--|



Workshop Agenda

22 September 2019

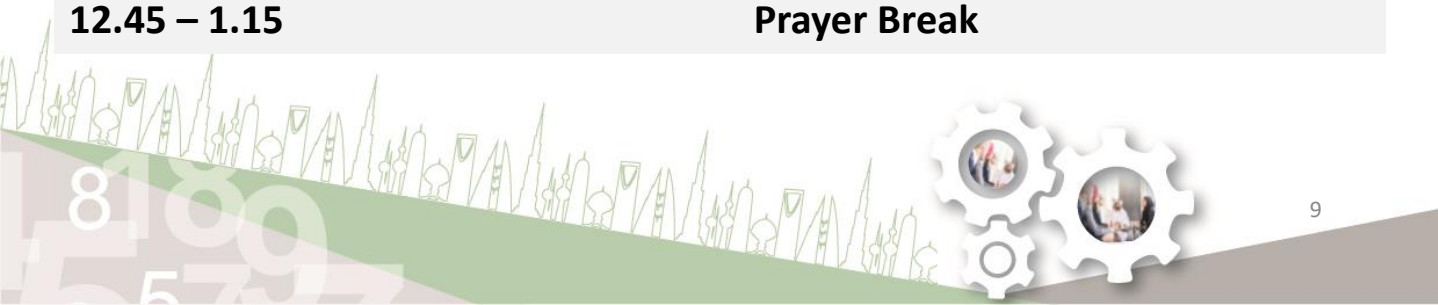
| Time | Speaker | Topic |
|--|------------------------|---|
| Session 2: Quality of Administrative Data | | |
| 12.00 -12.45 | Nancy McBeth | Applying GCC Quality Framework – Case Study 2020 Census -GCC Quality Framework -Quality control and Quality Assurance -Role of Editing in Quality Control and Quality Assurance |
| 12.45 – 1.15 | Prayer Break | |
| 1.15- 1.45 | Danilo Dolenc | Common Errors in Administrative Data – International Perspective - Types of errors in Administrative data – European Experiences - Common causes of error |
| 1.45 – 2.45 | Country Representative | Common Errors in Administrative Data – GCC Experiences - Country Presentations |
| 2.45 – 3.00 | Elham Salah | Review of Day 1 |
| Lunch | | |



Workshop Agenda

23 September 2019

| Time | Speaker | Topic |
|--|-------------------------|---|
| Session 3:Editing (Data Cleansing Techniques) | | |
| 8:30 – 9.15 | Danilo Dolenc | Errors in Identifiers and Linking Variables -Impact of errors in Identifiers and Linking variables -Identifying errors in Identifiers and Linking variables -Correction Techniques |
| 9.15 – 10.30 | Danilo Dolenc | Editing within Individual Sources -Identifying errors in Administrative Sources -Correcting errors – Principles and Practices -Special cases - Missing Records - Missing values Discussion |
| 10.30- 10.45 | | Coffee Break |
| 10.45 – 11.45 | Danilo Dolenc | Editing between different Register Sources - Identifying inconsistencies between different sources - Statistical approaches for managing inconsistencies Discussion |
| 11.45 -12.45 | Country Representatives | Experiences in GCC Countries Country Presentations |
| 12.45 – 1.15 | | Prayer Break |



Workshop Agenda – 23 September 2019

| Time | Speaker | Topic |
|---|---------------|--|
| Session 3: Editing (Data Cleansing Techniques) | | |
| 1.15- 1.45 | Nancy McBeth | <p>Using Survey Data to help Validate/Correct Administrative based Data</p> <ul style="list-style-type: none"> - Using Survey Data to check/validate patterns of administrative data - Can survey data be used to correct individual administrative data records? |
| 1.45 – 2.45 | Danilo Dolenc | <p>Using Administrative Data to help Validate/Correct Survey Data</p> <ul style="list-style-type: none"> -Using Administrative data to check/validate patterns in Survey data Using administrative data to correct/change individual survey responses |
| 2.45 – 3.00 | Batoul Obaid | Review of Day 2 |
| Lunch | | |



Workshop Agenda

24 September 2010

| Time | Speaker | Topic |
|---|---|--|
| Session 4: Implementing Editing Techniques | | |
| 8:30 – 9.00 | Nancy McBeth | Feedback to Agencies <ul style="list-style-type: none"> - Feedback on Quality - Maintaining support of Agencies - Managing differences in the administrative and statistical views of the data |
| 9.00 – 9.30 | Batoul Obaid | Documentation <ul style="list-style-type: none"> - Documenting editing rules and decisions - Metadata |
| 9.30 -10.30 | Danilo Dolenc | Editing in the Slovenian Census – Case Study <ul style="list-style-type: none"> - Overview of Slovenian Census - Role of Editing |
| 10.30- 10.45 | Coffee Break | |
| Session 5: Practical applications | | |
| 10.45 – 11.45 | All | Practical Session 1 – Identifiers and Linking Variables <ul style="list-style-type: none"> -Practical activities |
| 11.45 -12.45 | All | Practical Session 2 – Editing Within Sources <ul style="list-style-type: none"> -Practical activities |
| 12.45 – 1.15 | Prayer Break | |
| 1.15- 2.15 | All | Practical Session 3 Editing across Sources <ul style="list-style-type: none"> -Practical activities |
| 2.15 – 2.30 | Elham Salah Batoul Obaid Nancy McBeth | Review of Workshop |
| 2.30 – 3.00 | Closing Ceremony | |
| Lunch | | |

4 8 1 8 9
5 5 7 9
1 2 5 7 9
6 9 4
7 2 4
3 4



المركز الإحصائي
لدول مجلس التعاون لدول الخليج العربية
GCC-STAT



[http:// www.gccstat.org](http://www.gccstat.org)



<https://facebook.com/gccstat>



twitter.com/gccstat



info@gccstat.org



ص.ب: 840، الرمز البريدي: 133
سلطنة عمان



P.O.Box:840, PC:133
Sultanate of Oman

+ 968 24346499 :



+ 968 24343228 :

