

Quality assessment of multi-source statistical processes

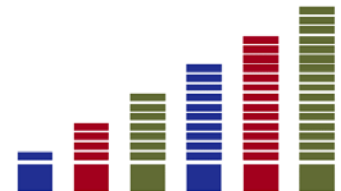
Session 18 - Multi-Source Statistics

Giovanna Brancato, Antonia Boggia, Andrea Bruni

Istat, Italian National Statistical Institute

Outline

- **Istat quality assessment Programme:** Brief illustration
- **Quality model** in statistical processes using administrative data
- The **assessment questionnaire:** main areas
- Final remarks and future work



Istat Quality Assessment Program



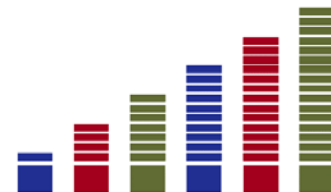
Direct assessment

Auditing and self-assessment programme on \approx 14 processes each year

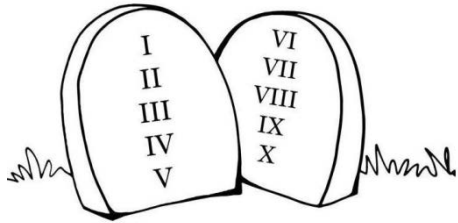


Indirect assessment

Analyses of Standard Quality Indicators stored in Istat Quality Documentation System (SIDI-SIQual)



Istat Quality Program: Direct Assessment



Principles in Quality Guidelines

Voluntary statistical processes

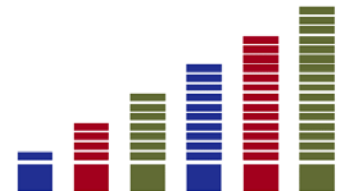
Auditing interview by
questionnaire



- Final assessment reports with improvement actions
- Report on the main findings



Follow-up of improvement actions

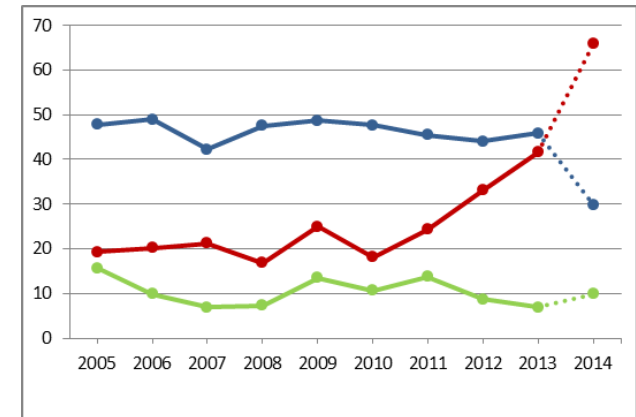


Istat Quality Program: Indirect Assessment

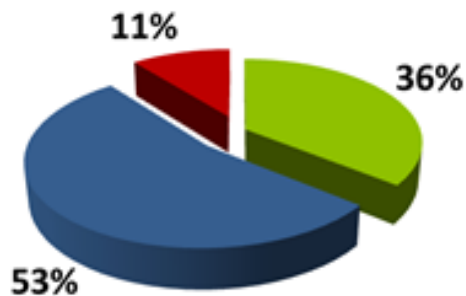
Examples

	Year 2011		Year 2010	
	mean	n	mean	n
ASIA	6,84	7	14,11	10

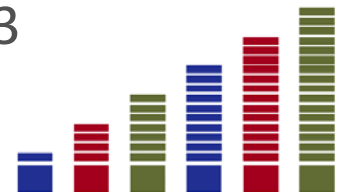
Mean weighted Frame Error Rates for Active Business Frame (ASIA)



Mean weighted Unit Nonresponse Rates for business (blue), households (red) and institutions (green) over time



% of yearly processes that have improved (green), maintained (blue) or worsened (red) timeliness in 2013 vs. 2012



Administrative source data: Quality model

Usability

Before any specific statistical purpose is identified

Input quality

Quality of the administrative sources centrally acquired

Input output oriented quality

Quality of the administrative data used in each statistical production process

**The Assessment
Questionnaire**

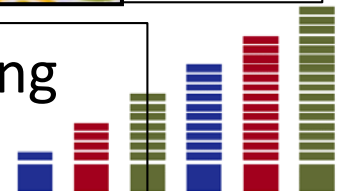
Through-put quality

Errors arising during integration of administrative data into statistical production process



Output quality

Quality of statistics produced using administrative data

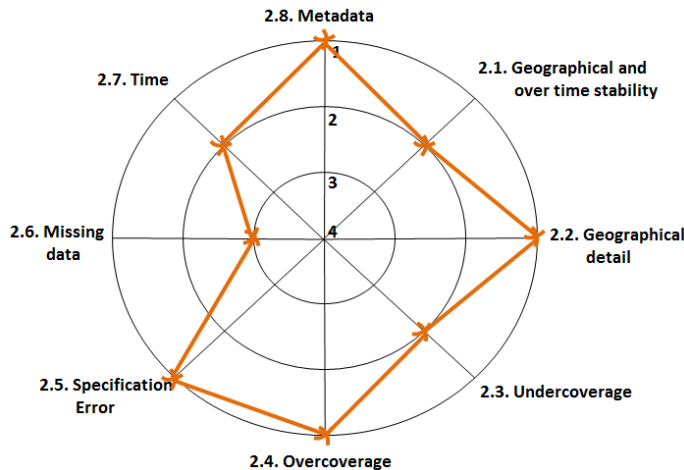
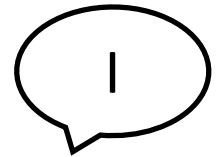


The Assessment Questionnaire

Input output oriented quality

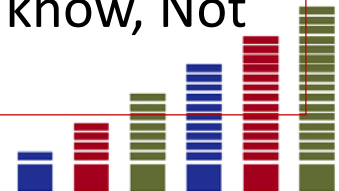
8 questions on the main dimensions of **data input quality (ORANGE QUESTIONS)**:

- 2.1. Geographical and over time stability
- 2.2. Geographical detail
- 2.3. Undercoverage; 2.4. Overcoverage
- 2.5. Specification error
- 2.6. Missing data
- 2.7. Time-related quality
- 2.8. Metadata



One final question on **overall assessment**

Possible answers: # datasets falling in:
 Excellent, Good, Fair, Poor, Don't know, Not applicable



The Assessment Questionnaire

Through-put quality

Specify Needs

1.1
Identify needs

1.2
Consult &
confirm needs

1.3
Establish output
objectives

1.4
Identify concepts

1.5
Check data
availability

1.6
Prepare business
case

Collect

4.1
Create frame & select
sample

4.2
Set up collection

4.3
Run collection

4.4
Finalise collection

Process

5.1
Integrate data

5.2
Classify &
code

5.3
Review &
validate

5.4
Edit & impute

5.5
Derive new
variables &
units

5.6
Calculate
weights

5.7
Calculate
aggregates

5.8
Finalise data
files

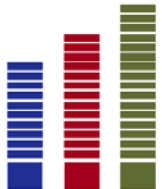
5.9.
Align time
references

The Assessment Questionnaire

Structure of the section

Example: Integrate Data

- Principle (*as stated in the Quality Guidelines*)
 - Questions on planning/re-planning
 - Questions on applied practices, methodologies & techniques, methodologic soundness
 - Questions on sources of error: e.g. errors in the linkage key and in the linkage procedure (linkage errors)
 - Questions on the documentation of the sub-process
 - Question on overall quality of the sub-process (process quality)



Example: Integrate Data

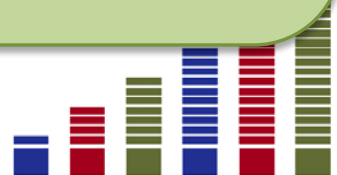
- Questions on sources of error:
 - ✓ Existence of measurements for the errors
 - ✓ Appraisal of the errors (GREEN QUESTIONS)

3.2.14. Assessment of false linkage error (based on indicators or not)

How do you appraise the quality of the linkage with respect to false linkage errors?

- 1 Excellent
- 2 Good
- 3 Fair
- 4 Poor

Specify:



Quality diagram: Source of errors

Sources of error considered:

Linkage variable quality
False link, False no-link

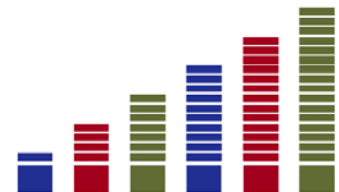
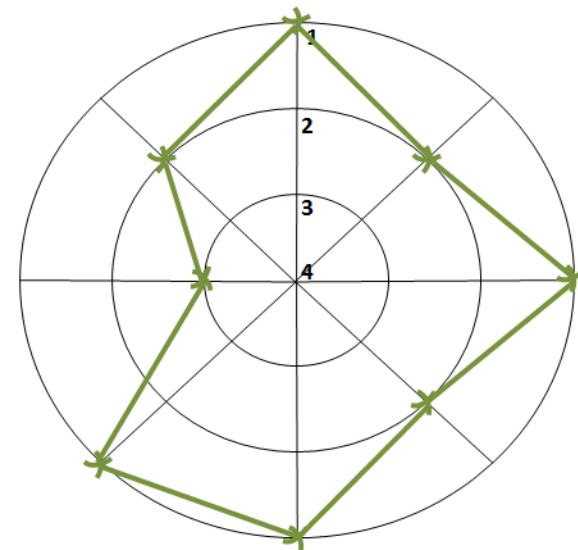
Population Overcoverage,
Undercoverage, Representativeness

Over time and geographical
comparability (variable)

Data completeness, measurement error

Model assumption error

Microdata timeliness and punctuality



Example: Integrate Data

- Question on overall quality of the sub-process (process quality) (BLUE QUESTION)

3.2.17. Quality of the Integrate data sub-process

How does the process manager/the team of auditing overall appraise the quality of the data integration procedure?

- 1 Excellent
- 2 Good
- 3 Fair
- 4 Poor

Specify:

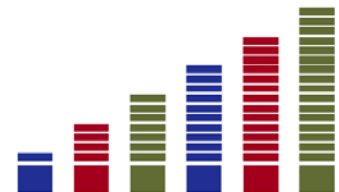
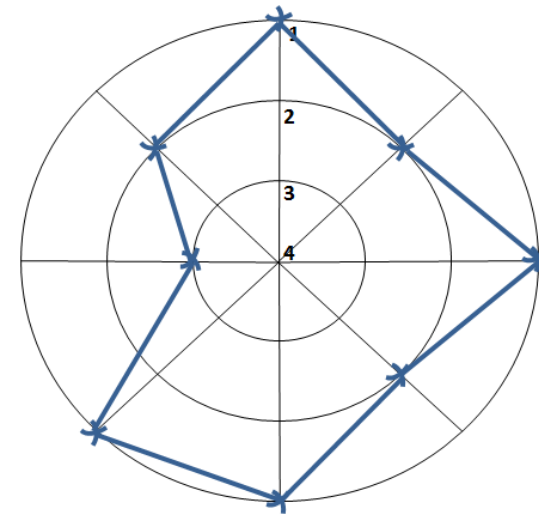
P



Quality diagram: Sub-processes quality

Process quality components considered:

- 3.1.11. Phase of identification of needs
- 3.2.17. Data Integration
- 3.3.9. Deriving units procedure
- 3.4.8. Deriving variables procedure
- 3.5.12. Time-related issues
- 3.5.13. Geographical-related issues
- 3.6.18. Editing and imputation procedure
- 3.7.9. Estimation
- 3.8.7. Validation procedures
- 3.9.12. Archiving, confidentiality,
dissemination and documentation



Quality diagram: Output quality

Output quality

Set of questions

- impact of the use of admin sources on each output quality dimensions
- appraisal on quality dimensions (GREY QUESTIONS)

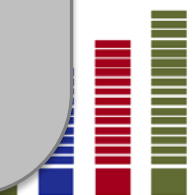
Example: Comparability over time

4.4.9. Comparability of estimates over time

How does the process manager/the auditing team appraise the comparability of the estimates over time?

- 1 Excellent
- 2 Good
- 3 Fair
- 4 Poor
- 5 Not applicable (no previous estimates)

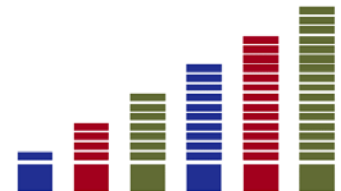
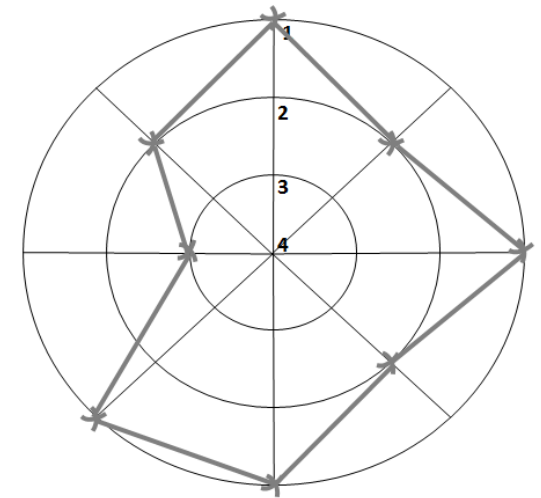
Specify:



Quality diagram: Output quality

Dimensions considered (Eurostat):

- 4.1.4. Relevance
- 4.2.5. Accuracy
- 4.2.12. Reliability
- 4.3.5. Timeliness
- 4.3.6. Punctuality
- 4.4.5. Coherence
- 4.4.9. Comparability
- 4.5.4. Accessibility and clarity



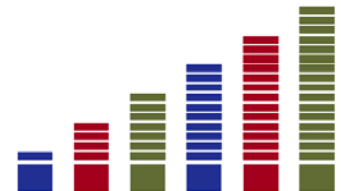
Final remarks and future work

Done

- Questionnaire for processes using administrative data
- Test on 3 processes
- Suitable in all situations
- Complexity of administration when high # of sources
- Design of the Questionnaire for mixed sources processes

Future work

- Test and fine tuning of mixed sources-questionnaire
- Set up indirect assessment
- Extension of the assessment program to any kind of statistical production



Thank you for your
attention

brancato@istat.it

