Quality Assurance: UNECE Activities

Steven Vale
UNECE
steven.vale@un.org
No specific framework …

- Around 70 member countries of Conference of European Statisticians
- EU countries use ESS QAF
- Others use NQAF
... but various quality activities

- Quality indicators for the GSBPM
- Quality of Big Data
- Quality in the context of data integration
- Quality chapters in various methodological guidelines

Likely future topics
- Quality in national data ecosystems
- Quality of geospatial information
Quality and the GSBPM

Documentation

- “The GSBPM provides a structure for organising and storing documentation within an organisation, promoting standardisation and the identification of good practices”
Quality and the GSBPM

Integrating metadata and quality

“The common framework provided by the GSBPM can help to integrate international work on statistical metadata with that on data quality by providing a common framework and common terminology to describe the statistical business process”
### Mapping SIMS to GSBPM

#### Eurostat Metadata Task Force

<table>
<thead>
<tr>
<th>Quality Management / Metadata Management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4</strong> Collect</td>
</tr>
<tr>
<td><strong>5</strong> Process</td>
</tr>
<tr>
<td><strong>6</strong> Analyse</td>
</tr>
<tr>
<td><strong>7</strong> Disseminate</td>
</tr>
</tbody>
</table>

#### 4.1 Create frame & select sample
- Sampling frame (S.21.1.1.1)
- Data collection (S.21.3)
- Sampling procedure (S.21.3.3)
- Quality assurance (S.13.1)
- Integration of new units and transactions (S.21.3.6)

#### 4.2 Set up collection
- Data collection (S.21.3)
- Frequency of data collection (S.21.2)

#### 4.3 Run collection
- Questionnaire (S.21.3.2)

#### 5.1 Integrate data
- Confidentiality - data treatment (S.8.2)
- Source data (S.21.1) - reuse
- Statistical survey (S.21.1) - use
- Administrative sources (S.21.1.2)
- Other sources (S.21.1.3)
- Priority for data sources (S.21.1.4)

#### 5.2 Classify and code
- Classification system (S.4.2)
- Checks of coding and data errors in source data (S.21.4.1)
- Integration of new units and transactions (S.21.3.6)

#### 5.3 Interpret & explain outputs

#### 6.1 Prepare draft outputs
- Documentation on methodology (S.12.1)
- Derogations (S.20.3)
- Adjustments (S.21.6)
- Seasonal adjustments (S.21.6.1)
- Quality documentation (S.12.2)

#### 6.2 Validate outputs
- Data validation (S.21.4)
- Administrative data checks (S.21.4.2)
- Quality documentation (S.12.2)

#### 6.3 Interpret & explain outputs

#### 7.1 Update output systems
- Quality documentation (S.12.2)
- Availability of data vintages of revisions (S.20.1.1)
- Data revision - practice (S.20.2)
- On-line database (S.11.3)

#### 7.2 Produce dissemination products
- News release (S.11.1)
- Publications (S.11.2)
- Other (S.11.5)
- Documentation on methodology (S.12.1)
- Inventories and manuals (S.12.1.2)
- Quality documentation (S.12.2)
Quality and the GSBPM

Process quality management

“If a benchmarking approach to process quality assessment is to be successful, it is necessary to standardise processes as much as possible. The GSBPM provides a mechanism to facilitate this”
GSBPM quality indicators

- Developed by task team
  - Canada, Hungary, Italy, Turkey, Eurostat, UNECE
- Generic quality indicators for each GSBPM sub-process
  - Processes based on surveys: 2016
  - Expanded to include administrative data: 2017
- Consistent with existing frameworks:
  - UN NQAF, ESS Code of Practice / QAF
Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources

(Version 2.0, October 2017)

https://statswiki.unece.org/display/GSBPM/Quality+Indicators
Uses of the Quality Indicators

- To provide a standard framework / common terminology to support a process-oriented approach to Quality Management
- To rationalise quality work within an NSO
- To define a mid-term quality policy
  - Set quality targets for a 3-5 year period
Example: 4.4 Finalise Collection

<table>
<thead>
<tr>
<th>Quality Dimension</th>
<th>Indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost-effectiveness</td>
<td>Discrepancy between planned versus actual collection costs</td>
</tr>
<tr>
<td></td>
<td>Percentage of collection activities that met requirements (assessed through analysis of paradata)</td>
</tr>
<tr>
<td>Accuracy and reliability</td>
<td>Outgoing error rates; estimate of non-sampling error</td>
</tr>
<tr>
<td>Accuracy and reliability</td>
<td>The rate of over-coverage: The proportion of units accessible via the frame that do not belong to the target population (are out-of-scope).</td>
</tr>
</tbody>
</table>
Quality of Big Data

- “A Suggested Framework for the Quality of Big Data” – December 2014
  - Australia, Canada, France, Italy, Mexico, Netherlands, Poland, Slovenia, UNSD, UNECE
- “extensions to existing statistical data quality frameworks were needed in order to encompass the quality of Big Data”

https://statswiki.unece.org/display/bigdata
Quality in data integration

- Part of UNECE guidelines prepared by international project
- Uses Zhang's two-phase life-cycle method model for integrated statistical microdata

https://statswiki.unece.org/display/DI/Quality