



HUNGARIAN
CENTRAL
STATISTICAL
OFFICE

Agenda Item 16

Data quality and metadata

Lidia Bassó

lidia.basso@ksh.hu

UNSD Seminar on International Trade Statistics
26-30 September 2011, Vilnius

Data Quality

Use of regular quality measurement

- **External**

- inform users on reliability of statistical data

*Think over carefully **what** and **how** to publish!*

- **Internal**

- compare indicators of subsequent periods
- identify areas where improvements are needed

EU quality reporting and assessment procedure

- Structure along EU quality dimensions (see later)
- Specific for ETS, separate chapters for Extrastat and Intrastat
- **National Quality Report of MSs** – *by MSs* (partially prefilled)
 - Q indicators (when data available in Eurostat; to be checked and missing filled in)
 - Info on methods & practices (last year's to be updated + expected changes)
- **Individual Assessment Report** – *by Eurostat against*
 - legal requirements
 - recommendations
- **Comment on assessment** – *by MSs*
- **Summary Assessment Report** – *by Eurostat*
Not mentioning MSs by name
- **Summary Quality Report** – *by Eurostat*
Each item of National Quality Reports of all MSs

Hungarian Quality Report

- *Along the same quality dimensions as EU QR*
- ***Common framework for all statistics***
 - select the stat. indicators at which Q is assessed (e.g. imports, exports, trade balance)
 - only the relevant Q indicators filled in
- ***Annual report***
- ***Gradual completing***
 - Q indicators for each dissemination
 - Data final – interpretation of indicators + plans for improvements
- ***Some points for consideration***
 - To assess improvement within domain – NOT to compare different domains
 - Careful consideration on when and what to disseminate to users
(only internal use now)

Quality dimension 1 – *Relevance*

- ***EU Quality Report***
 - Non satisfied user needs
 - Info required by EU legislation but not provided
- ***HU Quality Report***
 - Main users by groups (other dpts, national, EU, non-EU)
 - Requirements of above users and their legal basis
 - Measurement of user satisfaction

Quality dimension 2 – *Accuracy*

- ***EU Quality Report***
 - *Coverage* (shares of estimated values)
 - *Revisions* (rate of change in value) at aggregated level, detailed level, revision policy
 - *Confidentiality*
N° of CN8 codes affected, impact on value and net mass, conf. rules
 - *Control procedures*
description, share of electronic declarations
- ***HU Quality Report***
 - *Sampling errors*
 - *Non-sampling errors*
non-response, measurement, processing, coverage errors, effect of revisions

Quality dimension 3 – *Timeliness & Punctuality*

- ***EU Quality Report***
 - Time lag between end of ref. month and first transmission to Eurostat
 - Number of late transmissions to Eurostat, average delay
- ***HU Quality Report***
 - Time lag between end of ref. period and first/last dissemination
 - Delays to dissemination schedule

Quality dimension 4 – *Accessibility & Clarity*

- ***EU Quality Report***
 - National dissemination channels of EU data
 - Metadata to EU data disseminated nationally
- ***HU Quality Report***
 - Ways and channels of dissemination
 - Standard tables on Internet
 - Free Internet database
 - Other publications (download, paper, CD etc.)
 - Data transmissions
 - Safe room for researchers
 - On individual request

Quality dimension 5 – *Comparability*

- ***EU Quality Report***
 - *Over space*
asymmetries at aggregated and detailed levels + methodological reasons
 - *Over time*
changes in concepts and definitions with an impact on EU data
 - *Across domains*
methodological reasons for discrepancies
- ***HU Quality Report***
 - *Over space*
international rules/recommendations on concepts and definitions, compliance
 - *Over time*
since when data available, changes in past 15 years having impact on comparability (what and how), length of comparable time series, coherence between first and final figures
 - *Coherence with national accounts*
rate and reasons for discrepancy

Quality dimension 6 – *Performance, cost & response burden*

□ *EU Quality Report*

- *Values of thresholds*
- *Share of trade value in categories acc. to thresholds*
- *Share of enterprises in categories acc. to thresholds*
- *Simplification measures implemented*

□ *HU Quality Report*

- *Related costs for Statistics Hungary*
- *Burden on providers of stat. info (time spent + other costs)*

Metadata

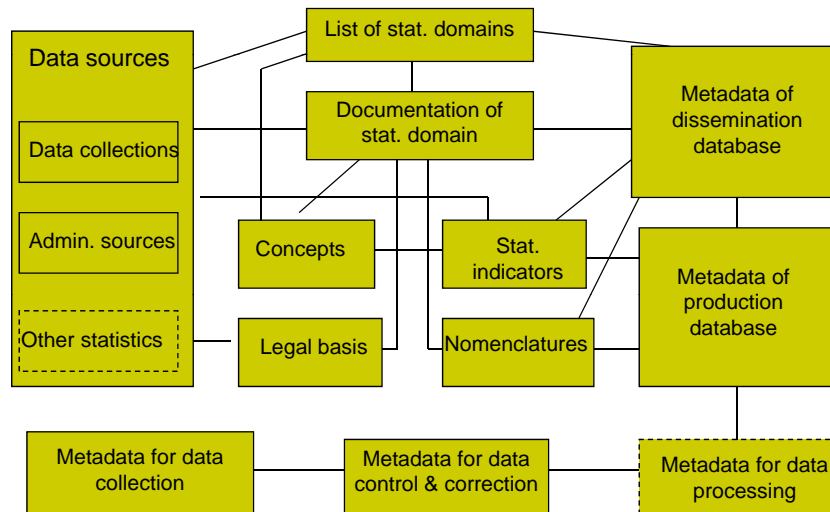
Statistical metadata

- *Description or numerical info on the statistical data by*
 - content
 - form
 - characteristics
 - sources
 - quality
 - methods used along the production
- *Some examples*
 - definition of a concept
 - items of a nomenclature
 - description of a method applied
 - identification of data providers for a data collection
 - by text for users
 - by algorithm for IT programmes
 - quality indicators

Metadata are used for ...

- *Informing users* on the source, characteristics and production process of data and methodology behind
 - basic descriptions easy to understand for non-educated, more detailed for educated users
- *Detailed documentation for statisticians*
 - complete description of methodology, practice, quality
- *Control info for the operation of programmes and softwares* of data collection, control & correction and processing, operation of data warehouse
- *Common basis for the integration of statistical systems*
 - e.g. nomenclatures, stat. indicators
- *Common standard for international data transmissions*

Schematic diagram of the meta system in Statistics Hungary



Thanks for your attention!

*What do YOU think about
data quality and metadata?*