Supervisory Data Quality Framework

Patrick Hogan
Head of Section, Supervisory Statistics Division

Supervisory Reporting Conference
28th November 2016

European Central Bank, Frankfurt
Overview

1. Introduction and characteristics
2. Data quality dimensions and metrics: Outcomes
Introduction: environment

<table>
<thead>
<tr>
<th>Governance and Infrastructure</th>
<th>Risk Data Aggregation Capabilities</th>
<th>Risk Reporting Practices</th>
<th>Regulatory Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Governance</td>
<td>• Accuracy and integrity</td>
<td>• Accuracy</td>
<td>• Review</td>
</tr>
<tr>
<td>• Data Infrastructure &amp; IT infrastructure</td>
<td>• Completeness</td>
<td>• Comprehensiveness</td>
<td>• Remedial actions and supervisory measures</td>
</tr>
<tr>
<td></td>
<td>• Timeliness</td>
<td>• Clarity and usefulness</td>
<td>• Home/host cooperation</td>
</tr>
<tr>
<td></td>
<td>• Adaptability</td>
<td>• Frequency</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Distribution</td>
<td></td>
</tr>
</tbody>
</table>

**ECB Data Quality Framework**

• Accuracy and Reliability
• Completeness
• Punctuality (timeliness)
• Consistency
• Plausibility
• Stability
Comprehensive approach to data quality

- Several approaches are taken to assess data quality of the received reports.

<table>
<thead>
<tr>
<th>Data quality</th>
<th>Missing data, delays and resubmissions</th>
<th>Plausibility of reported amounts</th>
<th>Number of data points, countries and currencies reported</th>
<th>Failing EBA validation rules</th>
<th>Basic internal consistency checks</th>
<th>Basic data points always reported</th>
</tr>
</thead>
</table>
The supervisory statistics data quality framework

• Applied and selected set of dimensions and metrics to frame the assessment of supervisory data quality within the Supervisory Statistics Division.

• Based on the ECB Statistics Quality Framework and inspired by other practices of data quality implemented among international institutions.

• Efficient framework applicable to different units of observation, different levels of aggregation and different time dimensions.
Definitions

Dimensions:

According to the Principle 2 of the UN Principles of Official Statistics (1994) states: “to retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.”

The list of these considerations responds to a set of dimensions which apply to supervisory statistics: punctuality, accuracy and reliability, consistency and plausibility.

Metrics:

• Metrics are define as system or standard of measurement: a set of figures or statistics that measure results.

• They have to take into account the nature and the type of data analysed

• They have to measure a specific dimension in an unbiased way, i.e. being interpretable and comparable across time and unit of observation (e.g. ratios, skewness of the distributions).

• They have to measure a specific dimension in an absolute way, i.e. making the user able to understand the actual magnitude of the comparison (e.g. levels, growth rates).
Framework adopted for Supervisory Statistics

Relevant Dimensions for Supervisory Data:
- Accuracy and Reliability
- Plausibility
- Consistency
- Punctuality

Units of observation:
- Report
- Template
- Data point

Aggregation levels:
- SSM-wide
- Country
- Peer group
- Institution

Time-frame:
- Quarter of reference
- Total in the past four quarters
- Average in the past four quarters

Metrics characteristics:
- Adaptation to the nature of the data (it’s supervisory data!)
- Provision of unbiased information
- Magnitude of the evaluations
Overview

1. Introduction and characteristics

2. Data quality dimensions and metrics: Outcomes
Punctuality

Punctuality refers to the lag in time between the remittance date and the actual submission of the data from the NCA to the ECB (ongoing discussion to receive data on the submission from institutions to NCAs). The following metrics will be implemented both for each time-frame level.

Average days of delay before receiving the first module
# Accurate and reliable data

**Accuracy** shall be defined as the absence of mistakes and exact correspondence of the reported values with the underlying concept for each data point. **Reliability** refers to the closeness of revised values of a specific data point to the initial value released. Supervisory data shall accurately and reliably assess the phenomenon they are intended to measure.

## Maximum available European Banking Authority’s Validation Rules by reporting period

<table>
<thead>
<tr>
<th>Quarter of Reference</th>
<th>Blocking Validation Rules</th>
<th>Non-blocking Validation Rules</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q4 2014</td>
<td>1338</td>
<td>455</td>
<td>1793</td>
</tr>
<tr>
<td>Q1 2015</td>
<td>1338</td>
<td>455</td>
<td>1793</td>
</tr>
<tr>
<td>Q2 2015</td>
<td>1314</td>
<td>620</td>
<td>1934</td>
</tr>
<tr>
<td>Q3 2015</td>
<td>1314</td>
<td>620</td>
<td>1934</td>
</tr>
<tr>
<td>Q4 2015</td>
<td>2088</td>
<td>718</td>
<td>2806</td>
</tr>
<tr>
<td>Q1 2016</td>
<td>2088</td>
<td>718</td>
<td>2806</td>
</tr>
<tr>
<td>Q2 2016</td>
<td>2088</td>
<td>718</td>
<td>2806</td>
</tr>
</tbody>
</table>

*The number of applicable VRs in the table do not take into account their applicability to the reported data.*
Accuracy and reliability

• Failing validation rules within a template:
In the supervisory statistics environment, we view accuracy as the closeness of the reported value to the underlying supervisory concept. Equivalently, it is the degree to which data correctly reflect the corresponding ITS concepts. It is quantified using the number of failing validation rules and completeness checks.

Failing EBA validation rules
Data quality dimensions and metrics: Outcomes

Accuracy and reliability

• Completeness:
  Is defined as the availability of the largest and most material subset of the required information.

Percentage of submitted data points

<table>
<thead>
<tr>
<th>Module</th>
<th>Quarterly reports</th>
<th>Semi-annual reports</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1 2016</td>
<td>Q2 2016</td>
</tr>
<tr>
<td>COREP</td>
<td>90%</td>
<td>93%</td>
</tr>
<tr>
<td>FINREP</td>
<td>79%</td>
<td>85%</td>
</tr>
<tr>
<td>AE</td>
<td>78%</td>
<td>81%</td>
</tr>
<tr>
<td>LCR</td>
<td>79%</td>
<td>79%</td>
</tr>
<tr>
<td>NSFR</td>
<td>87%</td>
<td>89%</td>
</tr>
<tr>
<td>Total average</td>
<td>83%</td>
<td>85%</td>
</tr>
</tbody>
</table>

The data used in the table above comes from a set of pre-defined data points that are considered essential by supervisors to complete key supervisory tasks and should be reported by all institutions independent of their size, business model or country of origin. However, because of differences due to business models making some data points redundant for that entity, achieving 100% is not possible and completion rates over 80% are considered as satisfactory.

• Resubmission studies:
  Based on the analysis of the difference between preliminary and revised reported values. All significant resubmissions (according to the guidelines provided by the resubmission note) are automatically detected and lead to further investigations.
Consistency is defined as the satisfaction of logical relations between different subsets of the data (i.e. across templates), their correspondence with the master data associated to the institution and their correspondence with other published data. The reported information shall be consistent over time, across datasets and comparable with external data.

• **Internal consistency:**
  Reported values are consistent and reconcilable across templates. Moreover, values are consistent with the information provided in the master data.

• **External consistency**
  Various checks should be carried out to assess the consistency of the data received with other datasets, published balance sheets and internal ECB Data.
Plausibility checks aim to detect outliers in the reported data. This is accomplished by reviewing the time series of the variable concerned with both a statistical approach and a business-based one. Values that markedly deviate from the usual pattern of the series are isolated and further analysed.

- **Stability:**
  Change in the total number of reported data points from period to period
Plausibility

- **Outlier analysis:**
  Outlying unit of observations are flagged and explanations are requested to the institutions via the NCAs

  We look at values with:
  - extremely high (or extremely negative) **growth rates**.
  - extremely high (or extremely negative) **levels**.

The table shows the number of outlying values (potential non-plausible values) selected for Q2 2016 reported data, based on a wider subset of data points considered key for supervisory activities. The table shows the findings after applying a hierarchy of prioritisation based on a combination of several outlying indicators.
Thank you: Questions or observations