

## Supervisory Data Quality Index for Supervised Entities

### Abstract

*High-quality data is the precondition for effective risk analysis that can enhance the efficacy of supervisory processes covering large and highly interconnected banking and finance landscapes. This article seeks to provide an insight into a supervisory data quality framework which is based on data quality dimensions and metrics. The framework has good scalability and adaptability to meet the needs of data quality assessment across supervised entities over time. This article also provides a methodology for an effective data quality assessment with a dynamic feedback mechanism based on data characteristics. Finally, based on this methodology, we propose a Supervisory Data Quality Index (sDQI) that can measure supervisory data quality objectively.*

### Introduction

Similar to many supervisory authorities and central banks the world over, supervising key financial institutions is one of the core functions of the Reserve Bank of India (RBI). Since 1995, vertical / entity-specific supervision has been supplemented with horizontal and risk-focused surveillance to monitor the financial health of supervised entities (SEs). This pronged approach helps identify SEs that show financial weakness and warrant supervisory attention while giving insights into the systemic risks.

Accordingly, the scale and scope of data collected have increased manifold<sup>1</sup>. The Supervisors observe that improving risk data aggregation capabilities and risk reporting practices remain a challenge for SEs, and supervisors would like to see more progress (Bank for International Settlements, 2013). In this direction, it becomes vital to measure and monitor the quality of the data continuously. The monitoring of reporting is being done through in-built checks for data within and across returns and manual exercises for outliers' detection. However, three objectives of supervisory data quality monitoring require the creation of an objective measure with a relevant tolerance threshold.

First, the quality of risk and financial data should be measured objectively based on the accuracy, timeliness, and completeness of the reports (McGilvray, 2010, 2021). Second, the measure should be such that it allows comparison across entities and over time (Leo L. P., Yang W. L, and Wang, 2002). Third, the data quality assessment should be communicated

---

<sup>1</sup> In terms of scale, the Department of Supervision collects 68,390 entity-level data items in addition to 138 small to large-sized datasets containing granular data items (customer/ counterparty -wise, etc.) through 88 returns submitted by Commercial/ Cooperative Banks, Non-banking Financial Companies and All India Financial Institutions at varying frequencies. In terms of scope, the collected data points range from figures in final accounts (e.g., Balance sheet, Profit & Loss etc) to risk data aggregated from underlying business verticals.

to the entity, indicating the supervisory expectations. This acts as a trigger for timely remedial action (Li, 2023).

Considering the objectives mentioned above and taking reference from the European Central Bank (ECB) data quality framework<sup>2</sup>, a framework is proposed for the comprehensive assessment of the quality of data submitted through supervisory returns. The proposed sDQI framework measures the data quality based on the accuracy of the data submitted, the timeliness of the submission, the completeness of the data points submitted and the consistency with which the entity has been submitting the quality data. The Master Direction – Reserve Bank of India (Filing of Supervisory Return) Directions – 2024 issued on February 27, 2024,<sup>3</sup> has included the Basel Committee on Banking Supervision (BCBS) Principles for effective risk data aggregation and risk reporting<sup>4</sup> (2013), aiming at “*improved risk data aggregation capabilities and risk reporting practices*” of SEs. The circular is motivated by Principle 13 of *ibid*, which states that “*Supervisors should have and use the appropriate tools and resources to require effective and timely remedial action by a bank to address deficiencies in its risk data aggregation capabilities and risk reporting practices*”.

The sDQI provides an objective metric for the assessment of the bank’s adherence to the above BCBS principles. The sDQI also allows consistent assessment of the SEs’ adherence over time and also allows comparison with peers.

The remainder of the article is organised as follows. Section II presents the cross-country experiences of banking regulators and supervisors on data quality indicators. Section III discusses the detailed methodology of the sDQI computation along with the rationale for the selection of the constituent parameters; Section IV deals with the classification of SEs based on the sDQI score, and Section V concludes.

## **II. Cross-Country Experiences**

Many central banks, such as the European Central Bank and Federal Reserve Bank of Kansas City, prepare such Data Quality indices (DQI) based on robust methodologies on recommended data quality assurance/ assessment frameworks of other regulators / international agencies<sup>5</sup>. The ECB’s DQI has been published on their website every quarter since 2014. The DQI report is published at an aggregate level, giving parameter-wise details, the distribution of scores and trends. The individual reporting entity’s DQI is shared with its supervisors and included in the Supervisory Review and Evaluation Process (SREP) assessment as part of the “Risk Infrastructure, Data and Reporting” chapter of the “Governance” element of the SREP. Supervisors share, at their discretion, a bank’s individual

---

<sup>2</sup>[https://www.bankingsupervision.europa.eu/framework/statistics/shared/pdf/ssm.explanatorynoteonaggregateddataqualitytables\\_201901.en.pdf](https://www.bankingsupervision.europa.eu/framework/statistics/shared/pdf/ssm.explanatorynoteonaggregateddataqualitytables_201901.en.pdf)

<sup>3</sup> <https://rbi.org.in/web/rbi/-/notifications/master-direction-reserve-bank-of-india-filing-of-supervisory-returns-directions-2024>

<sup>4</sup> <https://www.bis.org/publ/bcbs239.pdf>

<sup>5</sup> Data Quality Assessment Framework (DQAF) of IMF and National Quality Assurance Framework (NQAF) of the United Nations.

DQI with the reporting institution.<sup>6</sup> ECB monitors data through data quality indicators based on data quality scores of reporting entities. ECB has developed a Data Quality Indicator based on the metrics for Punctuality, Accuracy, Completeness, Stability, Plausibility and Reliability. The aggregate data quality scores based on the first four parameters are published in the public domain at a quarterly frequency.

The International Monetary Fund (IMF) assesses the quality of data submitted by various countries under the General Data Dissemination System (GDDS)/ Special Data Dissemination Standard (SDDS)/ SDDS Plus through its Data Quality Assessment Framework (DQAF) prepared by it in July 2003 and subsequently revised in May 2012. The DQAF, which is used for comprehensive assessments of countries' data quality, covers institutional environments, statistical processes, and characteristics of the statistical products. The DQAF has five main dimensions- (i) assurances of integrity, (ii) methodological soundness, (iii) accuracy and reliability, (iv) serviceability and (v) accessibility.

The United Nations (UN) also has a National Quality Assurance Framework (NQAF) template and guidelines. Some of the important principles include Relevance, Accuracy and reliability, Timeliness and punctuality, Accessibility and clarity, Coherence and compatibility. Countries are asked to either follow the UN's NQAF or prepare their own National Quality Assurance Frameworks for data quality.

The Federal Reserve Bank of Kansas City measures the data quality into two discrete sets of four measures. The first set of four objectively measured data quality dimensions covers Completeness, Consistency, Uniqueness and Validity. Another set of four subjectively measured data quality dimensions are Accuracy, Flexibility, Timeliness and Usability.

### III. Methodology

sDQI measures the quality of data on the four dimensions of **Accuracy**, **Completeness**, **Timeliness** and **Consistency** (ACTC). The methodology of quantifying the same on the four mentioned dimensions is given below.

- a) **Accuracy** – The dimension of Accuracy, as the name suggests, assigns scores based on the accuracy of the data submitted in each return by each SE. The accuracy assessment is based on quantifiable concerns, viz. within and across return validations, unexplained deviation from past levels and trends, comparison with published data, observations by supervisory teams and findings during analytical assessments.
- b) **Completeness** – The dimension of Completeness assigns scores based on the data submitted in each return and the number of returns submitted. The assessment is based on the hypothesis that the values of all components of a single datum are valid, if not, then its completeness is destroyed (Wang & Storey, 1995). The completeness is assessed

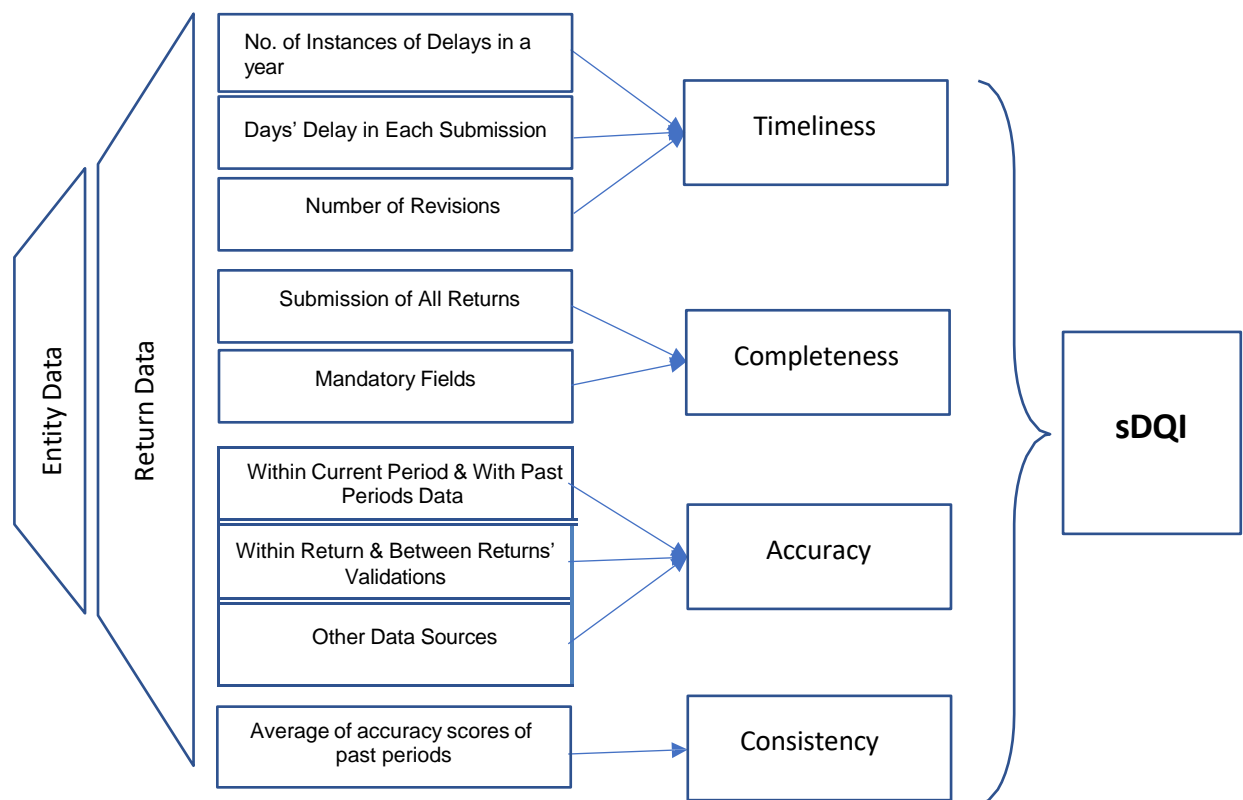
---

<sup>6</sup>[https://www.bankingsupervision.europa.eu/ecb/pub/pdf/annex/ssm.aggregateddataquality\\_first\\_quarter\\_2024\\_202406.en.pdf?d01a83bc31bba27b3b91351eff2a9f21](https://www.bankingsupervision.europa.eu/ecb/pub/pdf/annex/ssm.aggregateddataquality_first_quarter_2024_202406.en.pdf?d01a83bc31bba27b3b91351eff2a9f21)

based on return level and datum level submissions. The score for mandatory fields is assigned based on whether the SE has reported a zero value under such fields, which is expected to have a non-zero value.

- c) **Timeliness**: The Timeliness assessment is based on the timeline prescribed in Master Directions for the concerned return. The scoring is based on the instances of delay, the overall number of days of delay and the number of revisions for the past twelve months. The instances of delay are categorised into “without default”, “infrequent”, “More Frequent”, and “Repetitive” based on the Supervisory Action Framework.
- d) **Consistency** – The dimension of Consistency checks for the consistent accuracy of data for the previous quarters.

The entity-level sDQI score is arrived at by aggregating the scores for Accuracy, Completeness, Timeliness and Consistency. The scores are computed for the bank group (Public Sector Banks, Private Sector Banks, Foreign Banks and Small Finance Banks) and also for peer groups. The details of the above methodology are depicted in the following chart.



**Chart.1 sDQI methodology**

#### **IV. Classification of SEs based on sDQI score**

While relative comparison with peer group entities is computed, the sDQI is also rated in four bands based on the historical scores.

- **Good**: A score between 90 and 100 indicates that a SE has overall good compliance, indicating very few delays/errors in reporting and a sustained sDQI score.

- Acceptable: A score of between 80 and 90 indicates that the SE has a few reporting compliance issues that need to be addressed without undue delay, but the data reported is usable for supervisory examination.
- Needs Improvement: A score of less than 70 to 80 means that the SE has several reporting compliance issues that need to be addressed without undue delay, as it is not usable for supervisory examination.
- Major Concerns: A score of less than 70 means that the SE has serious issues in risk aggregation capabilities and requires supervisory intervention.

## V. Conclusion

The sDQI provides a comprehensive and quantitative measure of the supervisory data quality, forming the basis for supervisory examinations. The sDQI will help the supervisors assess the SEs in light of the guiding principles of BCBS, viz., accuracy, completeness, timeliness, comprehensiveness, consistency over time and comparability.

The publication of aggregated sDQI on the RBI website and the communication of entity-specific sDQI to the concerned SEs provide a standard method for assessment of SEs' data submission compliance. This is expected to bring better discipline in reporting supervisory data and also lead towards better inferences for the supervisors.

## References

1. Bank for International Settlements. (2013, January), "*Principles for effective risk data aggregation and risk reporting*", Bank for International Settlements, <https://www.bis.org/publ/bcbs239.pdf>
2. Li, M. L. (2023), "*Financial Data Quality Evaluation Method Based on Multiple Linear Regression*", Future Internet, 338.
3. McGilvray, D. (2010), "*Executing Data Quality Projects: Ten Steps to Quality Data and Trusted Information*", Beijing: Publishing House of Electronics Industry.
4. McGilvray, D. (2021), "*Executing Data Quality Projects*", Academic Press.
5. Leo L. Pipino, Yang W. Lee, and Richard Y. Wang. (2002), "*Data Quality Assessment. Communication of the ACM*", 45, 4 (2002), 211–218.
6. Wang, R., & Storey, V. (1995), "*Framework for Analysis of Quality Research*", IEEE Transactions on Knowledge and Data Engineering 1(4), pp 623–637.