**A Framework for Assessing the Quality of Banking Supervision Data**

Gregorio Guidi, ECB, [*Gregorio.Guidi@ecb.europa.eu*](mailto:Gregorio.Guidi@ecb.europa.eu)

Francesco Donat, ECB, [*Francesco.Donat@ecb.europa.eu*](mailto:Francesco.Donat@ecb.europa.eu)

Iulia Tintea, ECB, [*Iulia.Tintea@ecb.europa.eu*](mailto:Iulia.Tintea@ecb.europa.eu)

Luis Suarez Tumi, ECB, [*Luis.Suarez\_Tumi@ecb.europa.eu*](mailto:Luis.Suarez_Tumi@ecb.europa.eu)

**Abstract**

*Since the start of the economic and monetary Union, the European Central Bank (ECB) has placed strong emphasis on key aspects of statistical quality, as described in its Statistic Quality Framework. The introduction of a harmonised framework for supervisory reporting in 2014 has raised the awareness of competent authorities on the need to apply similar high standards of statistical quality to data collected for the purpose of banking supervision. Accordingly, the ECB has established a process for the quality assessment of supervisory data, with a twofold purpose. Firstly, checking whether data constitute a suitable basis for informing supervisory decision; secondly, recognising data quality as an integral part of banks’ supervisory evaluations. Following these two objectives, a number of data quality dimensions have been identified and implemented as part of a general framework for assessing the quality of banking supervision data. This paper presents some reflections on the framework, based on the experience gained at the ECB over the past four years.*

**Keywords:** Banking supervision; BCBS 239; Data quality; ECB; Supervisory reporting; Statistics Quality Framework

**1. Introduction**

The 2008 global financial crisis uncovered the vulnerabilities of the financial system and called for normative acts aimed at strengthening prudential regulation, supervision and cooperation at international level. Deficiencies in supervisory information available from the institutions, as well as fragmentation of national practices and lack of data harmonisation have been key areas of interest for the European legislators. Reliable data are in fact a prerequisite for banks’ risk monitoring as they allow for the prompt identification of potential weaknesses in the banking system and the coordination of timely actions between supervisory and resolution authorities (Rimmanen, 2015).

In this context, the European Banking Authority (EBA) introduced a mandatory harmonised reporting framework made up of two major schemes: a Common Reporting dedicated to the disclosure of regulatory requirements (COREP), and a comprehensive reporting of banks’ financial information (FINREP)[[1]](#footnote-1). This framework, implemented in 2014 for institutions in the European Union (EU), comprises a number of legal acts and technical documentation aimed at achieving uniform reporting via common definitions, templates and a Q&A mechanism managed by the EBA[[2]](#footnote-2).

As direct supervisor of the significant institutions of the countries participating to the Single Supervisory Mechanism (SSM), the European Central Bank (ECB) has been entrusted with the responsibility of ensuring the compliance of supervised institutions with the EU prudential rules, including supervisory reporting. In this regard, Article 140 of the SSM Framework Regulation[[3]](#footnote-3) entails the ECB to organise the processes relating to the “quality review of data reported by supervised entities subject to, and in compliance with, relevant Union law and EBA implementing technical standards”.

Given the relevance of quantitative information for decision-making, data quality has received a strong emphasis at the ECB. Under this respect, the *ECB Statistics Quality Framework* (SQF; ECB, 2008) recognises that “quality is interpreted in a broad sense as encompassing all aspects of how well statistical processes and output fulfil key stakeholders’ expectations” (p. 4). In other words it is expected that statistics within the ECB remain *fit for purpose*.

Despite this general definition, not all the principles set in the SQF are directly applicable to the ECB in the context of banking supervision data. The EBA retains the ownership of the reporting framework; therefore it contributes to the achievement of good supervisory data quality standards in conjunction with the ECB and the other national competent authorities (NCAs). By having the legal authority to propose amendments to the reporting framework, the EBA identifies *de facto* supervisors’ needs and translates them into specific data requirements from the institutions. The regulator ensures therefore the relevance and consistency of supervisory data[[4]](#footnote-4), as well as their timeliness and a basic level of accuracy by defining the data production cycle and the adherence of the remitted information to a set of established validation rules[[5]](#footnote-5).

If the role of the EBA in the quality of supervisory data is to ask the relevant questions (by providing consistent reporting templates), that of the competent authority is to check whether the institutions provide the correct answers (by filling the reporting templates appropriately). Following the work done on this topic by the ECB, the paper discusses some general principles underlying the quality assessment of supervisory data, and suggests a minimal set of dimensions that can help supervisors correctly frame the assessment.

**2. Principles for a Supervisory Data Quality Framework**

From the point of view of the supervisor-statistician in a competent authority, the quality assessment process starts the moment reporting templates are received from the institutions. Hidden from sight, however, are the many steps of the reporting process implemented by the supervised entity that contributed to generate the quantitative information submitted to the competent authority. From the initial inputs and transactions recorded by the banks, to the final reports, the path followed by the data is of staggering complexity. In fact, the process starts even before, with the interpretation of the prudential and reporting regulation, then passes through the definition of appropriate data structures and of IT solutions able to store and organise the granular information of the bank, and then proceeds in the application of complex transformation and aggregation rules to produce the high-level views that inform the banks’ management boards and fulfil the regulatory reporting obligations. This hidden complexity makes the role of the authority in charge of receiving and assessing supervisory data at the same time fundamental and challenging.

It is fundamental because deficiencies in the risk aggregation and reporting capabilities of a bank pose a serious threat to proper risk governance and risk-based decision making, for both the supervised institution and the supervisor.

It is challenging because it requires from statisticians faced with the task of assessing the quality of supervisory data to have a strong understanding of the supervisory process, of the applicable prudential and accounting frameworks, and of the banks’ business and internal organisation. In fact, “the data quality process requires a balance of statistical tools and supervisory expertise: statisticians and supervisors are both necessary for ensuring that the reporting is formally correct and accurate in substance, before being released and used” (Chionsini et al., 2017).

Automatic validation rules are the first line of defence in the quest for satisfactory data quality. It is however the role of the supervisor to set higher expectations and to take remedial actions if such expectations are not met.

In terms of supervisory expectations, at the international level the bar is set by the Basel Committee on Banking Supervision (BCBS) *Principles for effective risk data aggregation and risk reporting* (BCBS 239). Compliance with the BCBS principles, which has been considered one of the top priorities by the ECB in 2016 and 2017 (Pellizzari and Poloni, 2017), involves not only banks’ internal risk reports. Instead, it hinges on a holistic approach to reporting that encompasses also regulatory and external reporting.

In the context set by the BCBS 239 principles, the ECB approach to the quality assessment of supervisory data is twofold. On one side, supervisors want to know whether the data provided by the bank are a suitable basis to support and inform supervisory decisions. The assessment should then focus primarily on *completeness* of the data (meaning that all expected information should be available) and on its *accuracy*, to make sure that potential errors are flagged before they taint the decision-making process. On the other side, the data quality assessment is an integral part of the supervisory evaluation, providing hints of (non-)compliance with the BCBS 239 principles and uncovering issues in the reporting process, which could be symptoms of weaknesses in the banks’ risk management, systems and controls. Indeed, the assessment, summarized by a data quality score (see section 2.3), contributes directly to the Supervisory Review and Evaluation Process (SREP) – through the “internal governance and risk management” SREP element. For this part of the assessment, it is important not only to look at the latest available data submitted by an institution to highlight possible errors but also to enquire on how the information was provided: was it reported on time? Was it without errors from the beginning? Were values significantly revised without a clear explanation? Therefore, the analysis of completeness and accuracy should be complemented by a more comprehensive view of the reporting process. We argue that a meaningful off-site analysis should at a minimum investigate whether reporting templates are submitted on time (*punctuality* of the information) and whether later revisions stay close to the initial values reported (*reliability* of the information – following the terminology of the *ECB Statistics Quality Framework*).

*2.1 Assessing the quality of the data: completeness, accuracy and plausibility.*

Measuring *completeness* – that is, identifying missing modules or templates – is a relatively straightforward exercise once the reporting obligations have been determined.

On the other hand, *accuracy* (defined as the absence of mistakes and exact correspondence of the reported values with the underlying concept) is not so easily captured. To make it a practical dimension of analysis we rely mainly on validation rules: while they cannot guarantee that a reported value is *per se* correct, they can at least single out cases where there is a reasonable suspicion that the value is *not* correct. In this way, the analysis is able to produce *actionable findings*: specific instances of potential errors that are brought to the attention of the reporter, to be eventually corrected or explained.

In the context of checking accuracy, validation rules can be interpreted as *hard rules*, that is, rules for the *ex-ante* assessment of the compliance of the reporting entities with legal requirements and reflecting the basic constraints of the reporting templates. Ideally, the hard rules to measure accuracy applied by the competent authorities should be public, transparent and binding, helping reporters in the interpretation of the applicable regulations and of the structure of the reporting templates. But contributions to accuracy can also come from *soft rules*, interpreted as checks for the detection of situations that are unusual or unlikely to happen, serving as an *ex-post* mean for the assessment of the compliance with reporting obligations and also for the monitoring of banks’ activities.

Soft checks are used to challenge reporting institutions on the correctness of the data remitted to the ECB, following an *explain-or-resubmit* approach. It is in the supervisors’ interest to not disclose (some of) these checks to the reporters: as many soft checks rely to a certain extent on expert judgement and subjective thresholds, disclosing the checks can potentially convey the message that banks should comply with the threshold specified by the checks, distracting from the main objective of a general improvement in the quality of the reported data. The accuracy of the data is assessed on a continuous basis by the ECB, so that it will be unfeasible to determine aprioristically whether a submission will be fully correct or not. To this end, the approach taken at the ECB is to consider valid a submission without failing hard rules, implicitly deemed of a quality level sufficient to take unbiased supervisory decisions. Further analyses based on soft rules are usually left to a later stage.

This balance between accuracy and timeliness of information is enhanced by a continuous definition and application of new validation rules. While the bulk of validation rules are defined by the EBA, the ECB strives to cover additional parts of the reporting templates with new rules, developed in collaboration with the NCAs in response to specifically identified issues, or to reflect supervisory priorities. This process has been shown to be effective in bringing gradual data quality improvements.

Finally, the *plausibility* analysis, intended as the process to detect outliers in the reported data[[6]](#footnote-6), can also be conceptually associated with the soft validation rules. It is an important piece of the data quality puzzle, with the potential to provide additional actionable findings and help supervisors investigate and better understand the reasons for large variations observed in the data.

*2.2 Assessing the quality of the reporting process: punctuality and reliability*

Compared with the dimensions of analysis described in the previous section, punctuality and reliability provide indications on how the reporting process unfolded, without giving a point-in-time evaluation of the available data. *Punctuality* means that the reporting was submitted by the supervised entity to its supervisor within the prescribed remittance dates. There are only slight variations in the ways punctuality can be measured: it can be verified whether all data were available at the remittance date, or whether all data *without errors* were available at the remittance date, or how many days of delays passed until all data, or all data without errors, were made available[[7]](#footnote-7).

*Reliability* relates instead to the principle that values should not be significantly revised during a reporting cycle[[8]](#footnote-8). In the context of supervisory statistics, we can say that if revisions are such that they have the potential to change the assessment of the bank made by the supervisor, then reliability of the data is an issue. Making a detailed assessment of the potential effect of revisions is out of reach for the statistician, nonetheless we believe statisticians can provide a valuable service by highlighting to the supervisors those cases of frequent significant resubmissions and investigating their roots.

*2.3 The data quality assessment: from single findings to an overall measure of data quality.*

Data quality assessment is an integral part of the supervisory evaluation of a bank. As such, it is crucial that information on data quality is collected, summarized and communicated effectively to the supervisors.

In particular, the results of the data quality assessment should be made available at different levels of detail. At the higher level, it should be possible to capture the essence of the assessment in simple indicators or scores. This facilitates prioritisation and proportionality of supervision: more resources can be devoted to the cases where there are indications of data quality issues, to be further drilled-down to the lower levels (and eventually to the single issues), while the assessment could remain at the aggregate level otherwise.

Accordingly, we suggest that the assessment should be structured as a multi-layered approach. In a first layer, individual findings are communicated to the banks to receive explanations or resubmissions. Here, we emphasize the importance of a continuous feedback between supervisors and reporters to clarify and resolve in a timely manner all pending issues, which sometimes require several rounds of interactions[[9]](#footnote-9). In a second layer, aggregate measures are produced capturing a picture of data quality at the level of single reporting areas or dimensions. The following are examples of measures in current use: percentage of missing reporting modules or templates (completeness), number of failing validation rules and of open plausibility findings per module (accuracy and plausibility), compliance or non-compliance with the reporting deadlines and average days of delay (punctuality), number of significant resubmissions, number of data points added, removed or changed in resubmissions (reliability).

In a final layer, the data quality measures are summarized and condensed into a *data quality score*. The purpose of the score is to inform the dialogue between the supervisors and the bank in the context of the SREP, and to guide further investigations. Deficiencies identified during the data quality assessment process – summarized by a worse score – become then a valuable input to drill deeper into the bank’s risk data aggregation and reporting capabilities. Indeed, the recently released *Report on the Thematic Review on effective risk data aggregation and risk reporting* (ECB, 2018) shows how the level of implementation of the BCBS principles remains to this day unsatisfactory, noting that “weaknesses stem mainly from a lack of clarity regarding responsibility and accountability for data quality”.

**3. The Framework in Practice: Data Quality in Figures**

As an example of the application of the principles described in the previous sections, we show in the charts below two typical metrics useful to monitor general trends in the data quality of supervisory statistics, focusing on punctuality (Chart 3.1) and accuracy (Chart 3.2).

**Chart 3.1** Punctuality: percentage of reports received on time and average delay

Source: ECB calculations on the basis of submissions in scope of Regulation (EU) No 680/2014.

Notes: The timeliness is measured with respect to the ECB remittance deadline for each reference period (which differs from the deadline for primary reporting to the national authorities). Figures are computed using a constant sample of reporting institutions.

**Chart 3.2** Validation rule errors by reference period and report type

Source: ECB calculations on the basis of submissions supervised entities of NSFR, LE, LCR, FINREP, COREP and AE between 31/12/2014 and 31/12/2017.

Notes: The chart focuses on the sub-sample of EBA validation rules applicable to each reference period from Q4 2014 to Q4 2017. In particular, validation rules applicable to semi-annual templates are not included. Figures are computed as of one month after the ECB remittance deadline for each reference period. The following acronyms were used: NSFR, Net Stable Funding Ratio; LE, Large Exposures; LCR, Liquidity Coverage Ratio; FINREP, Financial Reporting; COREP, Common Reporting; AE: Asset Encumbrance.

Both charts display a pattern of data quality improvements over time, showing in particular a marked improvement during the first reporting periods. In our experience it is indeed common to witness an “adaptation period” in correspondence with the introduction of new reporting requirements. During this initial phase, reporting institutions adjust and streamline their processes to produce more consistent and accurate reporting.

**4. Conclusions**

In this paper, we presented a general framework for the quality assessment of supervisory data. We highlighted the twofold purpose of the assessment: first to check whether the data are suitable for informing supervisory decisions, and second to be an integral part of the supervisory evaluation, exposing deficiencies in the banks’ risk data aggregation and reporting capabilities and possible weaknesses in their risk management.

This led to the definition of a minimal set of dimensions to frame the data quality analysis (completeness, accuracy, plausibility, punctuality, reliability) and to a layered approach involving interactions with stakeholders at different levels. The dimensions recognise explicitly data quality as a multifaceted concept that needs to be investigated under a holistic approach, conciliating the different needs of the various stakeholders.

Some questions remain open, especially in relation to the harmonisation of procedures at the EU level. Topics to be further explored include a resubmission policy, the different roles of the ECB and of the NCAs in the data quality process, and the formal involvement of senior management and of the auditing departments in the reporting process. The future frameworks will also benefit from the inevitable introduction of new techniques and more advanced methods, such as approaches suitable for “big data”, made possible by the availability of more granular datasets.

***This paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.***

**References**

Chionsini G., Garcia L. and Quagliariello M. (2017), How far have we come on EU banking statistics? Are we there yet? In Eurostat, *Selected papers from the 2016 Conference of European Statistics Stakeholders*, pp. 40-49, Publications Office of the European Union: Luxembourg.

European Central Bank (2008), *ECB Statistics Quality Framework (SQF)*. Available at: <https://www.ecb.europa.eu/pub/pdf/other/ecbstatisticsqualityframework200804en.pdf>

European Central Bank (2018), *Report on the Thematic Review on effective risk data aggregation and risk reporting*. Available at: <https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.BCBS_239_report_201805.pdf>

Pellizzari G. and Poloni, P. (2017), Risk data aggregation and risk reporting: A regulatory perspective. In Haines P., *BCBS 239: Guiding principles for compliance*, pp. 13-32, Risk Books: London, UK.

Rimmanen M. (2015), The Single EU Supervisory Reporting. In Quagliariello M., *Europe’s New Supervisory Toolkit*, pp. 29-59, Risk Books: London, UK.

1. Commission Implementing Regulation (EU) No 680/2014. Earlier versions of the COREP and FINREP frameworks were in use since 2008 on a voluntary basis at national level. [↑](#footnote-ref-1)
2. See <https://www.eba.europa.eu/regulation-and-policy/supervisory-reporting>. [↑](#footnote-ref-2)
3. Regulation (EU) No 468/2014 of the European Central Bank. [↑](#footnote-ref-3)
4. Principles 9 and 11 of the *ECB Statistics Quality Framework*. [↑](#footnote-ref-4)
5. Principles 10 and 12 of the *ECB Statistics Quality Framework*. [↑](#footnote-ref-5)
6. Typical plausibility findings include values that changed unexpectedly in comparison to previous reporting periods or across peers, as well as unexpected changes in the number of templates and data points reported. [↑](#footnote-ref-6)
7. Measuring punctuality can be problematic because of the time lag between primary reporting (to the national authorities) and secondary reporting (to the ECB). [↑](#footnote-ref-7)
8. See Principle 10 of the *ECB Statistics Quality Framework*. [↑](#footnote-ref-8)
9. An open question concerns the expectation of resubmissions for errors found in past reporting periods. This is particularly relevant as the audience of supervisory data expands from just supervisors to other business areas, more reliant on consistent time series (e.g. financial stability, economic research). [↑](#footnote-ref-9)