RISK MANAGEMENT MAGAZINE

Vol. 21, Issue 1

January – April 2026

EXCERPT

Bridging RDARR and credit risk models: a data lineage-driven framework for sound data governance

Alessandro Di Maria, Vincenzo Frasca, Dario Girardi

https://www.aifirm.it/rivista/progetto-editoriale/

Bridging RDARR and credit risk models: a data lineage-driven framework for sound data governance

Di Maria Alessandro (UniCredit), Frasca Vincenzo (UniCredit), Girardi Dario (UniCredit)[1]

Corresponding Author: Di Maria Alessandro (alessandro.dimaria@unicredit.eu)

Article submitted to double-blind peer review, received on 6th February 2026 and accepted on 22nd April 2026

Abstract

The effective implementation of the Risk Data Aggregation and Risk Reporting (RDARR) principles introduced in 2013 by BCBS 239 continues to pose significant challenges for banking institutions, particularly in ensuring consistent interpretation and application across complex reporting domains. More than a decade later, the European Central Bank (ECB) has reinforced and broadened the scope of these principles through increasingly prescriptive supervisory expectations, following persistent implementation gaps. This paper aims to interpret the RDARR principles and, through a practical example, outline the changes and requirements necessary to achieve full compliance for credit risk models.

In doing so, this study first provides a mapping matrix to determine which credit risk models are RDARR-relevant, as this represents a necessary prerequisite for a consistent and risk-based application of the framework. Then, it focuses on Pillar I credit risk models, which are among the most advanced in terms of data quality and data architecture, owing to the regulatory frameworks established by the ECB and EBA. Building on this starting point, the paper analyses the main gaps in the application of RDARR principles and proposes a remediation framework for Pillar I models that attempts to mitigate the compliance burden often associated with RDARR. In particular, it leverages the introduction of an end-to-end data lineage concept as a key enabler to strengthen data governance and support the systematic integration of data quality controls aligned with ECB expectations. The proposed framework, developed for Pillar I models, serves as a benchmark for the broader set of credit risk models identified through the matrix, guiding their progressive alignment with RDARR principles.

Keywords: EU regulation, Data Governance, Data Quality, Credit Risk

JEL Classification: G18, G32, C45, C49

1. Introduction

The ability of financial institutions to manage and aggregate risk-related data in a consistent and reliable manner has become a fundamental prerequisite for sound decision-making and robust risk governance. Risk data aggregation and reporting (RDARR) not only underpins effective day-to-day management but also serves as the basis for strategic steering, regulatory compliance, and supervisory oversight. High-quality data enhance institutions' capacity to quantify exposures, monitor adherence to limits, and anticipate vulnerabilities under stress, while also enabling digitalisation, automation, and IT cost efficiency (European Central Bank, 2024).

Despite these clear benefits, recent supervisory assessments conducted by the European Central Bank (ECB) have demonstrated that structural weaknesses in RDARR remain widespread (European Central Bank, 2025). Since the 2008 financial crisis—when deficiencies in data accuracy, timeliness, and integrity amplified systemic vulnerabilities—regulators have prioritised the implementation of the Basel Committee's BCBS 239 principles (Basel Committee on Banking Supervision, 2013). Yet, more than a decade later, full adherence remains elusive. The ECB's thematic review on RDARR of 2016, subsequent on-site inspections, and the Supervisory Review and Evaluation Process (SREP) have consistently revealed shortcomings in governance, fragmented data architectures, extensive reliance on manual adjustments, and weak data quality controls (European Central Bank, 2018). In terms of timeliness of risk reporting, in many cases production times for monthly risk reports exceeded 40 working days, undermining institutions' ability to respond to evolving risks (European Central Bank, 2024). These findings highlight the slow progress of institutions in strengthening their RDARR compliance, with governance shortcomings and legacy IT infrastructures frequently cited as root causes. Many institutions continue to treat RDARR primarily as a compliance burden rather than as a strategic advantage, as highlighted in the latest ECB's newsletter on this topic (European Central Bank, 2025). The ECB's supervisory strategy for the coming years thus aims to enforce more substantial remediation

[1] The views and opinions expressed in this paper are solely those of the author(s) and do not necessarily represent the official stance of UniCredit.

efforts, stressing that adequate RDARR capabilities are not optional but a minimum expectation under European banking law and supervisory standards. Against this background, the publication of the ECB's *Guide on effective risk data aggregation and risk reporting* (European Central Bank, 2024) represents a fundamental step in clarifying regulatory expectations. The guide's purpose is "*to describe the practices which, in the ECB's view, are necessary from the perspective of RDARR to ensure effective processes are in place to identify, manage, monitor and report the risks supervised institutions are or might be exposed to*" (European Central Bank, 2024, p. 5). It identifies seven key supervisory principles that banks are expected to integrate in order to build a robust and resilient RDARR framework[2].

Several industry studies highlight the economic benefits of strong data governance and data quality frameworks. For example, banks with high-quality data can achieve a 5–6% increase in revenue and up to 20% higher profitability, while an effective data governance can reduce operational costs by 15–20% (McKinsey, 2025; Number Analytics, 2025). Similarly, banks with mature governance frameworks demonstrate significantly improved risk prediction accuracy and fewer unexpected losses, underscoring the direct link between robust RDARR practices and financial performance. Moreover, sound risk data management can lead to lower capital and liquidity buffers (due to data quality issues) and reduce costs related to data-related fines and data adjustment teams (Oliver Wyman, 2024). The SAS Institute (2024) emphasises that poor data quality remains the "Achilles' heel" of risk management, leading to costly reconciliation errors and undermining supervisory trust. In parallel, the digitalisation of finance, accelerated by artificial intelligence and advanced analytics, further raises the stakes: institutions with advanced data architectures gain significant advantages in risk monitoring, while those reliant on fragmented legacy systems face mounting vulnerabilities (Bank for International Settlements, 2024; Thite, 2025; Heß, 2025). Moreover, PwC (2025) highlights how the RDARR program can play a strategic role in enabling AI initiatives by improving data quality, governance, and infrastructure, and ensuring granular and reliable data that supports automation, emerging risk analysis, and model risk reduction. High-quality data allows to build robust AI models, minimising bias and inefficiencies, and delivering trustworthy outputs. Moreover, RDARR can strengthen governance and risk culture by integrating non-financial risks into decision frameworks and empowering competent oversight bodies. As a result, besides being a compliance effort, RDARR can act as a key enabler of innovation, better decision-making, and digital transformation.

The aim of this paper is to define a possible approach to achieving compliance with the RDARR principles for Internal Ratings-Based (IRB) Credit Risk Modelling framework. In this context, Internal Ratings-Based (IRB) credit risk models provide an ideal use case for operationalising RDARR principles. These models inherently generate, transform, and validate risk-relevant data through structured workflows, data pipelines, and embedded control mechanisms. This makes them a controlled environment in which RDARR expectations can be tested, assessed, and refined in practice. It is worth noting that Pillar I models have always been under the Supervisors' spotlight since the publication of BCBS 239, which laid the foundations for what later followed in the regulatory framework on data quality topics (European Banking Authority, 2016; European Commission, 2021). Over the years, supervised banks have aligned with these requirements, raising data quality standards within their organisations—standards that can be considered largely consistent with the data quality principles set out in the RDARR guidelines. Although achieving RDARR compliance across banking processes spans organisational, technological, and governance dimensions—and may significantly disrupt existing strategic and operational plans—the approach proposed in this paper is designed to minimise such disruption when the process under analysis is the IRB Credit Risk Modelling framework.

In pursuing compliance with RDARR principles within IRB Credit Risk Modelling framework, a key premise is that not all are intended to operate at the level of a specific process. For example, Principle 1 requires accountability of the management body for RDARR, Principle 6 requires sufficient timeliness of internal risk reporting, and Principle 7 demands the effective implementation of RDARR remediation programmes[3]. These principles are inherently cross-sectional and require institution-wide arrangements that extend beyond individual process pipelines and therefore cannot be meaningfully assessed on a single process level. Consequently, this paper identifies and isolates the principles that are most relevant at process level and, in particular, for credit risk models within the IRB framework. Specifically, among the RDARR principles, those most directly applicable to credit risk models are Principles 2, 3[4], 4, and 5, as they impose requirements that align closely with the scope, structure, data flows, and operational mechanics of the IRB modelling framework.

In response to contingent, compliance-driven needs arising from the practical application of RDARR principles, this paper proposes a methodology that positions data lineage—often the missing piece in IRB model frameworks—as the key enabler for fostering compliance with Principles 3, 4 and 5. The methodology is articulated as a sequence of replicable steps that can be applied, with proportionality, across different credit risk modelling environments.
In particular, it consists of: (i) a scoping step aimed at identifying RDARR-relevant credit risk models through a principle-based mapping matrix derived from Principle 2; (ii) a translation step whereby selected RDARR principles are operationalised via a

[2] A short description of the principles is provided in the Annex.
[3] See Annex for a more detailed description of these principles.
[4] For Principle 3, only subparagraph 1 will be covered as it can be considered process-specific. Subparagraphs 2, 3 and 4 relate to the roles of management body and central data governance function, internal validation and internal audit respectively. For the purpose of this paper – which adopts a process-driven perspective – central governance topics are out of scope.

structured mapping to DCAM capabilities, reducing interpretative discretion; (iii) an implementation step focused on the definition of layered data quality controls—technical, functional, model-specific and output-level—supported by a RACI-based governance framework; (iv) an architectural step centered on the design of end-to-end data lineage enriched with metadata, glossaries and taxonomies; and (v) a validation step addressing traceability and auditability through anomaly detection mechanisms and centralised registers. Together, these steps define a coherent methodological pathway that supports reproducibility, scalability and supervisory interpretability.

2. Sufficient scope of application for credit risk models

In line with the methodological approach outlined in the Introduction, the identification of RDARR-relevant credit risk models represents the first operational step of the proposed framework. This step is intentionally principle-based and proportional, translating the Principle 2 of the ECB Guide on RDARR (2024), "Sufficient scope of application" into a structured, yet flexible, scoping procedure. The output of this step is a mapping matrix that supports consistent prioritisation decisions, rather than prescribing exhaustive inclusion rules.

The objective of this section is therefore to outline the approach for identifying RDARR-relevant credit risk models and integrating them into the bank's RDARR plans and projects, ensuring full compliance with regulatory expectations.

According to this principle, institutions should establish a data governance framework that covers all material entities, risk types – including credit, market, liquidity, operational, and third-party risk – and the entire lifecycle of the data.

With respect to models, the scope of the data governance framework should include key internal risk management models, such as (but not limited to) Pillar I regulatory capital models (e.g. internal ratings-based (IRB) approaches for credit risk), Pillar II risk and capital models, and other key risk management models (such as IFRS9 collective provisioning models and value-at-risk models). This encompasses both input data for model development and resulting model outputs (e.g. exposure at default, probability of default, or loss-given-default estimates), which are essential for managing the institution's risks.

Furthermore, the scope should include relevant reports, including internal risk reports used for decision-making and strategic steering, financial reports externally published together with annual financial statements, and supervisory submissions such as FINREP and COREP templates, EBA and SSM stress test exercises, and Pillar 3 disclosures.

Based on the external regulatory expectations outlined above, we propose identifying the set of RDARR-relevant models through the following steps:

1. Identify the list of "RDARR-relevance criteria", against which assessing the relevance of credit risk models in terms of RDARR compliance. These criteria are derived from ECB Principle 2 and enriched to reflect the specific context of credit risk modelling.
2. Identify commonly used credit risk models within large commercial banks and classify them into "model families" based on their regulatory purpose or strategic importance – such as their role in provisioning, business support, credit origination, or credit monitoring.
3. Map each "model family" identified in step 2. to the set of "RDARR-relevance criteria" identified in step 1. If a "model family" meets any of these criteria, the entire family – and consequently all models within it – should be designated as RDARR-relevant.

The objective of step 1. is to define the list of "RDARR-relevance criteria" that will guide the identification of RDARR-relevant models. Starting from the ECB Principle 2 criteria, we expand and refine them for better alignment with credit risk modelling practices. Each criterion is further assigned a unique acronym for consistent reference throughout this section:

- Internal risk reports for decision-making and steering processes (which we label as "IRR"): this includes internal risk reports supporting executive and board-level decision-making and strategic steering, encompassing Risk Appetite Framework (RAF) indicators, key performance indicators (KPIs), and risk reports segmented by material risk types (e.g. credit, market, operational, and liquidity risks).
- Externally published financial reports and financial statements("FRFS"): this encompasses financial reports intended for external stakeholders, including audited annual financial statements, quarterly disclosures, and other regulatory filings that contain risk-relevant data. These reports are subject to public scrutiny and carry significant accountability implications.
- Supervisory reports submitted to supervisory and regulatory authorities ("SURE"): this encompasses periodic regulatory reports, such as FINREP, COREP, supervisory data collection (STE), EBA and SREP stress tests, and Pillar III disclosures.
- Key internal risk management models ("IRM"): this includes all internal models integral to the bank's risk management and regulatory capital framework. This spans models under Pillar I (e.g. IRB credit risk models), Pillar II (e.g. ICAAP frameworks), and other key models such as IFRS9 Expected Credit Loss Models for accounting and provisioning purposes

Having defined the "RDARR-relevance criteria" against which credit risk models will be mapped to assess their RDARR-relevance, we now proceed with step 2., identifying the model families which are expected to be commonly employed by large commercial banks. At a minimum, these include:

- *IRB models*: Pillar I models used for regulatory purposes, i.e. for the calculation of Risk Weighted Assets and regulatory capital ratios (e.g. rating systems used for IRB purposes).
- *IRB-Like models*: regulatory models that are either in the process of authorisation or developed to ensure the needed experience requirement for IRB Roll-Out.
- *Pillar II models*: models used for Pillar II regulatory purposes, such as the calculation of Economic Capital (e.g. credit risk satellite models, credit portfolio models, etc.)
- *IFRS9 models*: models used for accounting purposes to determine Expected Credit Loss (ECL) and Loan Loss Provisions (LLPs), including PD/LGD/EAD, Transfer Logic and behavioural IFRS9 models.
- *Managerial Models*: models covering non-IRB perimeters different from the ones covered by IRB-Like models. Their purpose is to support Economic Capital evaluation and credit business processes (e.g. perimeters under Permanent Partial Use and models decommissioned with the entry into force of Basel IV but still used for internal purposes). This family also includes models used only for business support, such as credit origination and monitoring, which do not contribute to other processes such as Economic Capital (e.g. Acceptance/ Transactional Scoring models, PSD2 models, Early Warning models).

The above-identified list of model families is designed to be as general as possible, ensuring applicability across commercial banks regardless of their specific characteristics. However, for some institutions, one or more model families may not be entirely relevant – for example, in the case of non-IRB banks, the first two families (IRB and IRB-Like) would not apply. This does not impair the validity of the proposed framework, as the absence of one model family does not affect the applicability of the others.

Having defined both the "RDARR-relevance criteria" and the list of model families, step 3. entails assessing each family for RDARR relevance. A model family should be flagged as RDARR-relevant if it meets at least one of the previously defined criteria. This assessment can be visualised using a matrix, where the rows represent the identified model families, and the columns represent the Data Criticality Criteria.

Model Family	IRR (Internal Risk Reporting)	FRFS (Financial Reports and Financial Statements)	SURE (Supervisory Reports)	IRM (Key Internal Risk Management Models)
IRB Models	✓		✓	✓
IRB-Like	✓			✓
Pillar II	✓		✓	✓
IFRS9 Models		✓	✓	
Managerial Models				✓

Table 1: mapping matrix for identifying RDARR-relevant credit risk models

As shown in the table above, all identified model families are considered potentially RDARR-relevant, as they meet at least one of the defined "RDARR-relevance criteria". It is important to note that the matrix presented is not intended to be universally prescriptive; rather, it reflects what is typically expected for large IRB-authorised commercial banks, where modelling frameworks are particularly complex and Pillar I and Pillar II models play a central role in supporting internal risk measurement and reporting. In other contexts – such as less significant institutions or organisations with less complex modelling frameworks – managerial models may instead assume a more prominent role. These models can support strategic activities and critical processes, including Internal Risk Reporting, which, in large and complex banks employing IRB models, is primarily supported by Pillar I and Pillar II models. However, this does not imply that RDARR principles are less relevant or that compliance is merely formalistic for non-IRB institutions. On the contrary, RDARR constitutes a structural requirement for all banks, irrespective of the regulatory approach adopted for credit risk. High-quality, traceable and reliable data are indeed fundamental prerequisites for sound risk management and effective internal risk reporting in any institution. Therefore, also non-IRB banks should be sufficiently equipped for complying with RDARR requirements, in a manner that is proportionate to their size, complexity and risk profile[5]. For these banks, the scalability of the framework proposed in this paper hinges first in the establishment of a robust data quality framework that achieves a degree of reliability comparable to that required for IRB models, and then in the development of the data lineage aspects described in the next section. Such a framework ensures that RDARR expectations

[5] Notice that the principle of proportionality in the application of ECB RDARR expectations is defined by the ECB Guide itself, which states that "*The Guide comprises the minimum supervisory expectations compiled by the ECB […]. The ECB intends to follow up on these expectations in its supervisory activities on a case-by-case basis, in line with the principle of proportionality*".

can be applied proportionately, yet meaningfully, thereby supporting improvements in the quality of management information and strengthening overall decision-making processes.

As we will show in the next section, we expect certain model categories – specifically Pillar I models – to be better aligned with RDARR principles, particularly regarding data quality requirements. This alignment is reinforced by requirements set out in the ECB Guide to Internal Models (2025), which already establishes detailed and rigorous expectations regarding data quality, thereby allowing for enhanced consistency with RDARR standards. Conversely, alignment with other, more innovative RDARR principles – such as data lineage – is expected to be currently less pronounced. These gaps can be progressively addressed by implementing the steps described in the next section.
For the other families of RDARR-relevant models, including those prevalent in non-IRB institutions, banks can expect to invest significant effort to align with the applicable RDARR expectations. This is because, unlike Pillar I models, the level of maturity and adherence to certain principles – such as data quality – is generally lower.
A separate mention deserves to be made regarding the Standardised Approach (SA) used for the calculation of regulatory capital requirements. Although this approach does not rely on internal models, its relevance has significantly increased following the introduction of Basel IV output floor requirements, which mandate that institutions' risk-weighted assets cannot fall below 72.5% of the corresponding Standardised Approach RWA. This regulatory constraint effectively requires banks – especially those applying the IRB Approach - to ensure that SA capital metrics are robust, traceable and comparable to their internally-modelled counterparts. This is particularly important because the Standardised Approach is entirely dependent on the accuracy, completeness and traceability of granular input data - such as counterparty classification, exposure type, collateral characteristics and external ratings - which directly determine the applicable risk weights. Any deficiency in these data elements can therefore materially affect the calculation of SA RWA and, consequently, the output floor benchmark.

Consequently, the data infrastructure supporting Standardised Approach calculations must also adhere to RDARR principles, despite not being formally classified as an internal model. Ensuring full RDARR compliance for SA processes is therefore essential to enable accurate benchmarking between IRB "advanced" calculations and their Standardised reference values and to support supervisory expectations regarding the consistency and reliability of output floor monitoring.

Before concluding this section, a brief consideration is warranted regarding climate risk, which represents one of the most challenging areas for data aggregation and reporting. Climate-related information is often affected by structural data quality issues, limited historical depth, and high variability driven by scenario-based projections. These challenges are particularly evident for Scope 3 emission data, where completeness and traceability remain limited due to heterogeneous data sources and supply-chain dependencies. Although a detailed treatment of climate risk is beyond the scope of this paper, the RDARR-oriented elements discussed—such as data lineage, metadata frameworks and structured data quality controls—offer a conceptual foundation that could support future developments in this area by enhancing transparency, semantic clarity and the evidencing of data provenance.

3. Data Lineage as element of Integrated Data Architecture

Building on the conclusions of section 2, this section analyses end-to-end data lineage as a practical cornerstone for operationalising the ECB's expectations on integrated data architecture (Principle 4) within IRB credit risk models. While IRB frameworks generally rely on comparatively mature, regulation-driven data quality controls (Principle 5), institutions often face structural challenges in demonstrating full, auditable traceability of risk data across their lifecycle - from source systems, through transformations and controls, to model outputs and downstream reporting. The ECB's 2024 RDARR Guide explicitly reinforces data lineage as a key evidencing mechanism for architecture, data quality, and governance outcomes, highlighting its role in bridging two domains that, although intrinsically linked, have frequently evolved separately in practice: data governance and data quality. Data governance has often been implemented through high-level, compliance-oriented policies that were insufficiently translated into binding, executable process requirements, management body awareness, and the involvement of internal control functions, weaknesses repeatedly noted by the ECB in relation to BCBS 239 implementation[6].
By contrast, data quality controls in the IRB context have a longer and more established history, rooted in Basel IRB requirements and subsequent supervisory guidance, including the European Commission Delegated Regulation (CDR) 2022/439 on IRB assessment methodology (European Commission, 2021). In particular, the CDR embeds binding supervisory expectations on core data quality dimensions—such as accuracy, integrity, completeness, consistency and timeliness—which substantially reflect the key principles introduced by BCBS 239 and later reiterated within the RDARR framework. However, these requirements have often developed as stand-alone arrangements within the modelling process, primarily focused on parameter estimation and validation needs[7].

[6] ECB Report on the Thematic Review on effective risk data aggregation and risk reporting (2018).
[7] Governance arrangements have also historically been present within the more mature IRB model landscape. However, these arrangements were often designed and implemented within the modelling process itself, without necessarily aligning with the broader data governance principles established at the bank level.

As a result, IRB models have evolved with localised data quality evidence that is not fully embedded within a broader RDARR governance framework run at bank level. The ECB's renewed expectations make clear that these existing IRB data quality frameworks must be integrated into the institution's overall data governance model, with data lineage providing the operational means to connect governance and quality by enabling transparent tracking of data origins, transformations (rules), applied controls, and qualitative and quantitative attributes across processes.

From an RDARR implementation perspective, this creates a pragmatic pathway for IRB models. In a typical enterprise rollout, the central data governance function oversees process owners as they identify in-scope processes, perform process mapping, and—supported by the data owner—implement data lineage. The resulting lineage is then used to define data perimeters and determine where data quality controls should be placed. For IRB models, however, several prerequisites are often already in place due to their long-standing regulatory and validation history: process knowledge is more structured, and a sound data quality control framework—fully or largely aligned with Principle 5—is typically already operating. Implementing lineage in this highly mature data quality context therefore delivers mutual benefits. First, it allows existing controls to be anchored to end-to-end traceability, strengthening evidencing and auditability in line with Principle 4. Second, it can enhance the traditional IRB set-up by making ownership boundaries explicit, formalising roles and responsibilities, and strengthening supporting governance tools such as metadata repositories, business glossaries/data dictionaries, and anomaly registers—ultimately improving standardisation, transparency, and the overall effectiveness of the data quality framework itself.

It is worth noting that this IRB-anchored design also enables the broader scalability of the proposed methodology. By construction, the methodological sequence outlined in this paper does not rely on IRB-specific artefacts, but on general architectural components—lineage, metadata, ownership allocation and layered data quality controls—that can be progressively extended to modelling domains with lower initial maturity. In practice, the same steps applied to IRB models can be replicated for IFRS 9, stress testing or managerial models, with the only differentiating factor being the degree of baseline data quality and process formalisation already in place. The methodology that will be described can also be used as a way to progressively raise maturity of models outside IRB domain: data lineage helps map fragmented processes, metadata gives a clear and consistent meaning to variables sourced from different systems, and the layering of controls provides a structured way to align these models with RDARR expectations.

In this sense, IRB models are not a special case, but simply the starting point: once the methodology is applied in a mature environment, it can be scaled to other model families identified under Principle 2 in a proportionate manner. This allows institutions to reduce the overall compliance burden while moving towards a more consistent RDARR framework across all modelling areas.

Consistently with the proposed methodology, this study adopts the Data Management Capability Assessment Model (DCAM)[8] - developed by the Enterprise Data Management Council (EDMC) - as a neutral translation layer to support the interpretation of RDARR principles. It is not employed as a compliance benchmark or maturity assessment tool, but as an industry-recognised taxonomy that enables the systematic mapping of supervisory expectations to data management capabilities, thereby reducing subjectivity and ensuring methodological transparency.

By mapping the relevant RDARR principles to the corresponding DCAM components enables the derivation of a set of verifiable criteria that a Pillar I credit risk model would be expected to meet to demonstrate alignment with RDARR model-level requirements. The analysis begins with a gap-assessment exercise in which selected RDARR sub-topics are systematically associated with the most pertinent DCAM components, providing a structured basis for deriving practical guidelines to support the interpretation of RDARR expectations in the context of Internal Ratings-Based (IRB) credit-risk models. Table 2 reports a summary of the gap analysis of selected RDARR principles declined for IRB credit risk models and evaluated via DCAM[9]. The strategy would be starting from *Not Compliant* on Principle 4 (Integrated Data Architecture, Data Lineage) to gain benefits on all the other not *Fully Compliant* topics to demonstrate how the design and implementation of data lineage may represent the foundation for full compliance with the selected principles in the context of credit risk models and which are the key questions that such activity would raise. Indeed, starting from the definition of data lineage given by the ECB[10] the full implementation of the data lineage ensures, at the same time, the coverage of the selected principles.

[8] DCAM provides a structured reference model that articulates a set of capabilities commonly associated with mature data management practices. According to the EDMC technical documentation, the model defines the scope of organisational, architectural and operational requirements necessary to evaluate data accuracy, integrity, completeness and quality within a formalised control environment, and it is designed to support the examination of data management processes across complex organisational landscapes (EDMCouncil 2023).

[9] Note that the degree of compliance scores have been stated according to the assumptions described in this work and are illustrative. Compliance degree may vary from institution to institution.

[10] "*Data Lineage is the information about the movement and transformation of data from front (capture) to end and enables a bank to (i) understand if data quality controls are sufficient and well placed in the data flow, (ii) identify interconnections between data definitions and taxonomies, (iii) ensure that when data fields are loaded or transformed across or within systems they are still in line with the reporting requirements and definitions, (iv) support the identification of data points needed for specific ad hoc reporting needs, (v) in case of data quality incidents be able to track back the source of the issue in a timely manner and to (vi) allow traceability for (external) validation*" (ECB, 2024).

RDARR principle	Selected RDARR sub-topic	DCAM	Compliance degree (assumed)
Principle 3	Governance of DQ control framework	7.1.2 Accountable parties have been identified and roles and responsibilities have been assigned	Partially Compliant
Principle 4	Identification of CDE and data lineage design	4.5.2. Critical Data Elements (CDEs) have been identified and inventoried	Not Compliant
	Contribute to the definition of process relevant metadata (semantics, taxonomies and business terms)	5.2.1. Attribute level "business" definitions are defined, documented and approved by relevant stakeholders	Not Compliant
		5.2.2. Taxonomies and ontologies are created, documented, maintained and governed	
		5.2.3. Metadata is defined	
Principle 5	Controls definition on CDE	7.2.1. All relevant data have been identified and prioritized.	Partially compliant
		7.3.1. Data Quality 'control points' are in place along the full spectrum of the data supply chain.	
	Coverage of main DQ dimensions	7.2.2. Data is profiled, analysed and graded	Fully compliant
	Root cause analysis, escalation process and traceability	7.2.3. Data remediation has been planned, prioritized and actioned.	Partially compliant
		7.3.2. Data Quality Metrics are captured, reported and used to drive data remediation.	
		7.3.3. Root Cause analysis is performed	
		7.3.4. Data Quality processes are audited	

Table 2: Outcomes of the gap analysis between a hypothetical IRB model and RDARR principles, evaluated via DCAM matrix for

3.1 Addressing Principle 3 and strengthening Principle 5: why assigning roles and responsibilities is essential to data lineage implementation and strengthens the pre-existing data quality framework

While data lineage is often approached as a technical or documentation exercise, it can only fulfil its role under RDARR if it is embedded within a clearly defined governance framework. Without explicit ownership and accountability, lineage remains static and informational, rather than operational. From a methodological standpoint, data lineage implementation is conceived as a logical sequence rather than a technical prescription. The sequence starts from role assignment and ownership definition, proceeds through the identification of Critical Data Elements and the placement of data quality controls, and culminates in

traceability and auditability mechanisms. This sequencing is intended to guide implementation choices while leaving room for institution-specific solutions.

This section therefore shows how the assignment of clear roles and responsibilities constitutes a prerequisite for effective data lineage, making its implementation a governance challenge and, at the same time, enabling compliance with Principle 3 at process level and strengthening the foundations of a comprehensive and effective Data Quality framework. To make the connection between data linage and assignment of roles and responsibilities even clearer, it might be useful starting from the definition reported by the ECB Guidelines on RDARR principles (ECB, 2024), where **data lineage** is defined as "*information about the movement and transformation of data from its initial capture to its final destination*."

A fundamental design question, especially when dealing with models, is therefore where lineage starts and where it ends. In practice, lineage should start at the authoritative capture/source - including the first controlled ingestion point - and end at the risk artefact that is actually consumed for decision-making and reporting (e.g. model outputs used in capital calculation and downstream risk reports).

From an operational perspective, governance mechanisms supporting data lineage must ensure:

- Clear accountability for the design, maintenance, and evidencing of lineage across the data lifecycle.
- Consistency between the process view (how risk data are produced and consumed), the data/functional view (CDEs, quality rules), and the technical view (systems, ETL, metadata).
- Ownership and placement of data quality controls at the right nodes of the lineage, with measurable thresholds and escalation.

In light of such multiple features, no single role can credibly own lineage end-to-end in isolation. From here, follows that, assuming that the analysis is conducted at a level where each process produces a single metric (e.g. the final PD parameter used for capital requirements calculation), the data flow can be segmented at least between the following stakeholders:

- **Process Owner** – accountable for the risk metrics outcome (process view)
- **Data Owner** – responsible for defining the functional data flow and the production of the risk metrics (functional/data view)
- **Application Owner** – responsible of the technological solution supporting the production of the risk metrics (data ingestion, ETL)

Ideally, roles and responsibilities should be formalised through a RACI matrix[11], which can be thought as follows:

Element	Process Owner	Data Owner	Application Owner
Risk metrics production	A	R	I
Process mapping	A	R	I
CDE identification	A	R	C
Data Lineage design and implementation	C	A/R	C
Technical controls	I	C	A/R
Functional controls	I	A/R	C
Model-specific controls	C	A/R	I
Output reconciliations	A	R	C/I

Table 3: RACI Matrix for E2E IRB modelling framework process under RDARR

[11] Reference on RACI matrix and players e.g. Responsible, Accountable, Consulted, Informed. See University of Oxford – RACI responsibility matrix for details.

As the Process Owner remains accountable for the **(a.)**, (**b.**) and **(c.)** with Data Owner acting as executor (responsible)[12], the main advantage of this configuration lies in leveraging the Data Owner's expertise to establish a centralized view, minimizing the number of data layers that constitute the data lineage (c.)[13]. A practical way forward could be to proceed by business data domain[14] thereby reducing the joint effort required. In addition to functional and risk considerations, the Application Owner may act as a third contributor, consultable for clearing the technical dimension (i.e. physical data and tables).

Once roles and responsibilities are clearly defined and assigned, the Data Owner can act as a pivotal figure by **(d.)** identifying all Critical Data Elements (CDEs)[15] already covered by pre-the existing data quality framework and tracing them back within the first data layer or legacy sources (e.g. balance sheet data, external credit bureau data, internal credit data), where technical controls designed by the Application Owner are typically applied. This approach should ensure an E2E coverage of the whole data flow feeding the model. Subsequently, the coverage of the first data layer—and its related CDEs—relies on two integrated sets of controls:

1. Technical controls **(e.)** - Application Owner may include – for instance – integrity of the data (check number of records from month to month, check missing values), uniqueness (no duplicate records), technical domains (values of fields must be included in predefined ranges).
2. Functional controls **(f.)** - Data Owner may leverage on risk logics: stability of distribution of raw input data needed for the calculation of the model's indicators, trend analysis, consistency of key values among different tables.

The key benefit of this approach is that the first data layer is common to most models, allowing this activity to be executed once for all models.

After designing this common layer, **(g.)** Model-specific controls can be developed by the Data Owner with support from the Process Owner. This involves identifying, for each model or rating calculation step, the relevant CDEs (e.g. risk scores associated with single modules that compose the model). These controls are expected to focus on risk considerations, such as detecting shifts in variable distributions or calibration segments over time (e.g. using empirical distribution analysis or PSI).

Finally, the last set of controls – **(h.)** Output reconciliations - targets the accuracy and consistency of metrics reported to Senior Management. Here, the focus shifts to the model outputs and their evolution over time. This stage, considered a monitoring phase, is shared by the Process Owner and Internal Validation, emphasising high-level business and risk perspectives rather than data errors (which were addressed earlier). Reconciliations with other risk reports using the same outputs may also be required.

In conclusion, the operating model described above—consistent with DCAM expectations—provides a structured way to link data lineage design, CDE identification, and layered control implementation to process-level governance of data quality. On this basis, the gap analysis results presented earlier can be updated to reflect the improved coverage of the relevant RDARR sub-topics. Such update is reported in the table below:

RDARR principle	Selected RDARR sub-topic	DCAM	Compliance degree (assumed)	Compliance degree (achieved)
Principle 3	Governance of DQ control framework	7.1.2 Accountable parties have been identified and roles and responsibilities have been assigned	Partially Compliant	Fully Compliant
Principle 4	Identification of CDE and data lineage design	4.5.2. Critical Data Elements (CDEs) have been	Not Compliant	Fully Compliant

[12] From ECB 2024 "*Data owners responsible for key risk indicators and critical data elements throughout the complete aggregation process (front to end)*".

[13] "Data layers" is here meant as a data domain feeding a particular step of risk metrics production. In this use case, one may think the first data layer as the collection of raw data (e.g. personal data, financial statements, external information from credit bureau etc.).

[14] For example: the Data Owner knows that different rating models are fed with financial indicators coming from the same data source. The calculation logics of such indicators (i.e. which data that need to be aggregated) is the same across all models. The Data Owner may then decide to propose a centralized view of this process step that optimizes the number of nodes which must be represented through the data lineage.

[15] As defined in ECB 2024, Critical Data Elements are "those data elements that are used to calculate the key risk indicators and have a direct or significant impact on the value of the indicator or technical routine of the calculation and the reporting".

		identified and inventoried		
	Contribute to the definition of process relevant metadata (semantics, taxonomies and business terms)	5.2.1. Attribute level "business" definitions are defined, documented and approved by relevant stakeholders	Not Compliant	
		5.2.2. Taxonomies and ontologies are created, documented, maintained and governed		
		5.2.3. Metadata is defined		
Principle 5	Controls definition on CDE	7.2.1. All relevant data have been identified and prioritized.	Partially compliant	Fully Compliant
		7.3.1. Data Quality 'control points' are in place along the full spectrum of the data supply chain.		Fully Compliant
	Coverage of main DQ dimensions	7.2.2. Data is profiled, analysed and graded	Fully compliant	Fully Compliant
	Root cause analysis, escalation process and traceability	7.2.3. Data remediation has been planned, prioritized and actioned.	Partially compliant	
		7.3.2. Data Quality Metrics are captured, reported and used to drive data remediation.		
		7.3.3. Root Cause analysis is performed		
		7.3.4. Data Quality processes are audited		

Table 4: Outcomes of the gap analysis between a hypothetical IRB model and RDARR principles, evaluated via DCAM - UPDATED

3.2 Complement the data lineage to fully comply with Principle 4: Introducing Metadata, Glossary and Taxonomies

Once roles and responsibilities for data lineage have been established, the next step is to ensure that the data lineage itself is enriched with a coherent semantic layer. End-to-end traceability alone describes how data move and transform across systems and processes, but does not fully explain what the data represent, how they should be interpreted, and under which business and regulatory assumptions they can be used. For this reason, achieving full compliance with Principle 4 requires complementing data lineage with a structured metadata framework encompassing data definitions, glossaries, and taxonomies.

This section therefore focuses on how Critical Data Elements (CDEs) identified through data lineage can be anchored to their underlying metadata, ensuring consistency between technical data, business meaning, and regulatory definitions[16]. It introduces the logical building blocks of this semantic framework—data elements, attributes, business terms, data dictionaries, and taxonomies—and explains how they interact to support both horizontal traceability across processes and vertical alignment from business concepts to technical implementations. By doing so, the section clarifies how metadata act as the missing link between lineage – built on CDE - and data quality – built on the technical fields (Technical Data Element, TDE), enabling effective governance of CDEs and reinforcing RDARR expectations on integrated data architecture and data quality standards.

Indeed, as reported in the ECB guidelines on RDARR program, the data lineage is expected to be designed "on data attribute level" which is something different to (critical) data element:

> *[…] a data element contains information as an independent field while a data attribute, in general, is a single value description (i.e. its metadata, such as a business description of the content, type, format, etc.) for a data element (or data point or data object). As an example, data attributes are often stored as a column in a table and are used in the technical mapping to calculate key risk indicators, whereas data elements impact the specific indicator values.*

The excerpt reported above highlights a clear breakdown between what should be considered a metadata and what should be considered a (critical) data element. According to this definition, we can interpret a metadata as the collection of attributes linked to a particular CDE. As an example, imagine a table storing information about credit exposures in a bank. One of the columns is called "Exposure Amount". The value in a specific row of the "Exposure Amount" column — for example, €1,000,000 — is a data element. It is an individual piece of information: a specific field containing data used in risk calculations.

On the other hand, the data attribute (metadata) describing that column — for instance:

- Name: Exposure Amount
- Type: Numeric
- ...
- Unit: Euro
- Format: Decimal with 2 digits
- Business Description: Total outstanding exposure of a counterparty at reporting date
  — these are data attributes. They describe what the data element is and how it should be interpreted, stored, or processed.

In the calculation of a hypothetical Key Risk Indicator (KRI) (e.g. Large Exposure Ratio), or in any weighting procedure which involves client exposures, the actual €1,000,000 is used in the formula — that is the data element affecting the outcome. Meanwhile, the data attributes help ensure the correct data is used — they guide the mapping and validation (e.g. ensuring the number is in euros, not dollars, and represents the correct concept).

Having established these definitions, as explicitly stated in Principle 4, banks are expected to have in place an internal Glossary which collects all the high-level descriptions - which can be easily understood by any stakeholder - of the Critical Data Element that constitute the data lineage: these descriptions are called Business Terms (BT). This capability is also remarked by DCAM which claims that "Attribute level "business" definitions are defined, documented and approved by relevant stakeholders".

On a different level, as required by DCAM, we have ontologies and taxonomies, which, respectively, represent how data entities are logically involved in the process (e.g. "EAD is calculated from Exposure Amount") and how they are related (e.g. EAD at counterparty level → EAD at facility level). In the context of credit risk models, according to the complexity of the organisation and processes in scope, taxonomies can make use of prefixes or suffixes to contextualize the Business Data Element (BDE)[17] within the process. This approach can be particularly relevant when defining the data lineage, as it ensures alignment between identified Critical Data Elements, their granularity, and the underlying business definitions. For example, consider the Exposure Amount defined earlier: this represents a Business Term included in the glossary and recognized by all stakeholders with its own definition. If we were mapping a credit risk model process, we would likely refer to the BDE "Exposure at Default (EAD)". However, in this context, using "Exposure Amount", even though it reflects the same broad business concept, could be misleading or less precise, as multiple types of exposures exist. This implies that BDEs, through the presence of taxonomies, can be nested to ensure accuracy of representation and effectiveness of the underlying semantics. In our example, BDEs might include "Regulatory Exposure Amount" and "Accounting (IFRS9) Exposure Amount".

[16] From ECB 2025 "*The management of data taxonomies should entail complete and up-to-date data lineages on data attribute level (starting from data capture and including extraction, transformation and loading) for the risk indicators, and their critical data elements, identified as being within the scope of application.*"

[17] The .the meaning and role of that specific data within the process.

Alongside with the semantics (Business Term) and business (Business Data Element) dimension, the CDE can be seen from another perspective: the technical dimension - Technical Data Element, TDE - which represents the actual physical asset used within the process which undergoes the actual data quality controls.

In conclusion, a (Critical) Data Element should be understood as a three-layer construct in which a Business Data Element becomes "critical" when it is relevant for the calculation of risk metrics or regulatory outputs, and is therefore coupled with its semantic attributes (BT, BDE, taxonomies) and its corresponding technical implementation (TDE).

The collection of BTs forms the Glossary, while BDEs and TDEs together constitute the Data Dictionary—an asset that associates technical data elements with business descriptions relevant to a specific process. The distinction between Glossary and Data Dictionary lies in their purpose: the Glossary is semantics-driven, enabling stakeholders to understand the meaning of a Business Term within the entity's context; the Data Dictionary is functionality-driven, featuring data elements used in specific processes and describing their functional dimensions (e.g. which calculation they feed).

Who compiles these definitions?

- TDE: Application Owner provides all technical attributes.
- BDE: Process Owner provides business definitions; Data Owner defines rules and active data quality controls, completing—along with Application Owner's input—the information required for the Data Dictionary.
- BT: All stakeholders, including those using the same TDE/BDE in other processes.

In summary, complementing data lineage with a structured metadata framework allows institutions to move from purely technical traceability to semantic traceability, ensuring that risk data are not only traceable, but also correctly interpreted and consistently used across processes, models, and reports. By anchoring Critical Data Elements to business terms, taxonomies, and technical implementations, banks can achieve vertical alignment between business meaning and physical data, and horizontal consistency across the end-to-end data lifecycle.
This semantic enrichment is a necessary condition for fully complying with Principle 4, as it enables integrated data architecture to be both transparent and intelligible, and provides the foundation for effective data quality governance under Principle 5. Moreover, it prepares the ground for traceability and auditability requirements by ensuring that each data point can be unambiguously reconstructed, understood, and validated in its regulatory and business context, as discussed in the following sectin.
The outcomes in terms of RDARR compliance achieved after implementing the steps described in this section are reported in the table below:

RDARR principle	Selected RDARR sub-topic	DCAM	Compliance degree (assumed)	Compliance degree (achieved)
Principle 3	Governance of DQ control framework	7.1.2 Accountable parties have been identified and roles and responsibilities have been assigned	Partially Compliant	Fully Compliant
Principle 4	Identification of CDE and data lineage design	4.5.2. Critical Data Elements (CDEs) have been identified and inventoried	Not Compliant	Fully Compliant
	Contribute to the definition of process relevant metadata (semantics, taxonomies and business terms)	5.2.1. Attribute level "business" definitions are defined, documented and approved by relevant stakeholders	Not Compliant	Fully Compliant
		5.2.2. Taxonomies and ontologies are created,		Fully Compliant

		documented, maintained and governed		
		5.2.3. Metadata is defined		Fully Compliant
Principle 5	Controls definition on CDE	7.2.1. All relevant data have been identified and prioritized.	Partially compliant	Fully Compliant
		7.3.1. Data Quality 'control points' are in place along the full spectrum of the data supply chain.		Fully Compliant
	Coverage of main DQ dimensions	7.2.2. Data is profiled, analyzed and graded	Fully compliant	Fully Compliant
	Root cause analysis, escalation process and traceability	7.2.3. Data remediation has been planned, prioritized and actioned.	Partially compliant	
		7.3.2. Data Quality Metrics are captured, reported and used to drive data remediation.		
		7.3.3. Root Cause analysis is performed		
		7.3.4. Data Quality processes are audited		

Table 5: Outcomes of the gap analysis between a hypothetical IRB model and RDARR principles, evaluated via DCAM - UPDATED

4. Traceability and Auditability

Having established the governance framework supporting data lineage and the semantic structures required to interpret Critical Data Elements consistently, the final step is to ensure that these arrangements translate into effective traceability and auditability, as explicitly required by Principle 5. From a supervisory perspective, data lineage, metadata and data quality controls deliver value only if they enable institutions to reconstruct end-to-end data flows in a timely, granular and reliable manner and to demonstrate the soundness of such flows to validators, internal audit and supervisory authorities. In this context, traceability refers to the institution's ability to identify – through the data lineage - the origin of the data quality issue, understand the root cause and activate – if needed - the related escalation process. Auditability extends this concept by requiring that such reconstruction is not merely theoretical, but supported by consistent documentation, control evidence and data governance tools that allow independent review and challenge. This section therefore focuses on how end-to-end data lineage, enriched with metadata and embedded within a data quality framework, enables institutions to meet ECB expectations on traceability and auditability, with particular attention to the implementation of a central anomalies register[18].

The ECB guidelines (ECB, 2024) claim that data lineage must allow institutions both to track back the source of data quality incidents in a timely manner and to provide sufficient transparency to external validators. This dual function translates into a real expectation for auditors and validators: not only should there be a documented flow of data from origination to reporting, but this documentation should be sufficiently granular to enable the analysis of any root-cause underpinning data quality

[18] *"An up-to-date and complete overview ("register") of data quality issues and limitations, including (i) an assessment of the severity of these issues, (ii) a root cause analysis, (iii) a quantitative impact analysis of material/severe data errors on the risk and business areas affected, (iv) clearly defined processes and responsibilities for remediating and escalating data quality issues, depending on the materiality of the issues, (v) deadlines for remediation, and (vi) a date for effective remediation (including appropriate evidence)" (ECB, 2024).*

problems. In practical terms, areas of data lineage that remain abstract or incomplete—e.g. documenting only systems but not transformation rules—compromise both remediation and assurance. Additionally, the ECB's approach requires institutions to maintain central anomalies register. This register serves as the inventory of all identified data quality issues, capturing information such as anomaly type, severity, business impact, root cause, responsible owner, and resolution timeline.

For validators and auditors, the register might represent a crucial control instrument, as it provides structured evidence that can be reconciled against both data lineage and data governance responsibilities. By systematically recording incidents, the register ensures that anomalies are not treated as isolated events but as part of a controlled feedback loop in the bank's data quality management framework. On top of that, auditors and validators can select samples of anomalies, verify whether they have been correctly classified, assess whether remediation has been carried out within agreed timelines, and test whether the incident can be traced back through the lineage to its origin. More importantly, the completeness of the register itself is an auditable element, since ensures compliance with ECB guidelines and a solid defence tool within the wider data governance framework.

To operationalise these tests, auditors and validators can define ad-hoc indicators that measure the effectiveness of the anomalies register and its integration with the data lineage. For instance, one could define the following metrics:

- **Data Lineage-to-anomaly traceability ratio**, defined as the proportion of anomalies in the register for which the root cause can be fully reconstructed using lineage documentation. For example, if 120 anomalies are registered over a quarter and only 96 can be traced back through to their source, the traceability ratio is 80%.
- **Ratio between the total number of Critical Data Element identified and the total number of data elements within a specific process.**
- **Average time used for fixing issues,** better if broken down by incident severity.
- A complementary measure, the **recurrence index,** can be used to detect whether the same anomaly reappears across multiple cycles, suggesting that root causes are treated symptomatically rather than structurally.

However, it is crucial to remember that these KPIs, while quantitative, cannot replace a qualitative assessment of the anomalies register. Validators and auditors must evaluate whether the register is comprehensive, whether responsibilities for logging and resolution are clearly assigned, and whether updates are embedded in change management processes. In addition, the register should be readily available for supervisory review, ensuring that external validators can trace anomalies seamlessly through the data lineage. A register that is technically complete but poorly governed will fail the ECB's standard of being both traceable and auditable. In sum, the ECB's 2024 emphasis on traceability and auditability elevates the anomalies register from a mere register of data issues to a core data governance tool that underpins data quality, strengthens supervisory dialogue, and positions internal audit as a central actor in the institution's data governance maturity. The table reported below presents the final outcomes achieved after having implemented all the methodological steps.

RDARR principle	Selected RDARR sub-topic	DCAM	Compliance degree (assumed)	Compliance degree (achieved)
Principle 3	Governance of DQ control framework	7.1.2 Accountable parties have been identified and roles and responsibilities have been assigned	Partially Compliant	Fully Compliant
Principle 4	Identification of CDE and data lineage design	4.5.2. Critical Data Elements (CDEs) have been identified and inventoried	Not Compliant	Fully Compliant
	Contribute to the definition of process relevant metadata (semantics, taxonomies and business terms)	5.2.1. Attribute level "business" definitions are defined, documented and approved by relevant stakeholders	Not Compliant	Fully Compliant

		5.2.2. Taxonomies and ontologies are created, documented, maintained and governed		Fully Compliant
		5.2.3. Metadata is defined		Fully Compliant
Principle 5	Controls definition on CDE	7.2.1. All relevant data have been identified and prioritized.	Partially compliant	Fully Compliant
		7.3.1. Data Quality 'control points' are in place along the full spectrum of the data supply chain.		Fully Compliant
	Coverage of main DQ dimensions	7.2.2. Data is profiled, analyzed and graded	Fully compliant	Fully Compliant
	Root cause analysis, escalation process and traceability	7.2.3. Data remediation has been planned, prioritized and actioned.	Partially compliant	Fully Compliant
		7.3.2. Data Quality Metrics are captured, reported and used to drive data remediation.		Fully Compliant
		7.3.3. Root Cause analysis is performed		Fully Compliant
		7.3.4. Data Quality processes are audited		Fully Compliant

Table 6: Outcomes of the gap analysis between a hypothetical IRB model and RDARR principles, evaluated via DCAM – UPDATED

4.1 Interaction with the Model Risk Management Framework

While the framework proposed in this paper does not aim to intervene directly in Model Risk Management (MRM) processes, the evidencing produced through RDARR operationalisation naturally interfaces with several elements of the MRM lifecycle. In particular, end-to-end data lineage enhances transparency on data sourcing and transformations, which are often among the most resource-intensive elements of independent validation. Likewise, layered data quality evidence may support assessments of model performance, stability and potential sources of model risk, without introducing RDARR-specific triggers or decision rules. Furthermore, anomaly tracking and ownership attribution strengthen documentation available for model change governance, again without redefining MRM practices, which remain governed by existing regulatory frameworks.

Importantly, the objective of this paper is not to integrate RDARR into the MRM framework nor to propose new validation methodologies. Instead, the intention is to highlight that the operational artefacts generated by RDARR implementation—data lineage, governance roles, metadata, and anomaly registers—can act as a complementary evidencing layer that supports existing validation and governance activities while remaining fully within current supervisory expectations.

A related regulatory development worth noting is the EBA's 2026 Final Report on the revised Regulatory Technical Standards for material model changes, issued on 30 March 2026 (EBA 2026), which streamlines supervisory approvals and strengthens documentation expectations for IRB model changes (e.g., clarity of data dependencies, consistency of supporting evidence, documentation of impacts). Although these RTS do not modify RDARR obligations, they underline the supervisory value of structured, transparent and traceable data processes—elements that RDARR implementation provides by design. The

connection is therefore indirect but meaningful: RDARR artefacts can facilitate compliance with evolving documentation expectations for model changes, easing the activities needed to be carried out by modelers, validators and auditors when interfacing with the regulator.

5. Conclusions

This paper addressed the persistent gaps in banks' Risk Data Aggregation and Risk Reporting (RDARR) capabilities by translating the ECB's 2024 principles into a practical, process-level remediation strategy. After fulfilling Principle 2 through a risk-based scoping methodology to identify RDARR-relevant credit risk models, the study demonstrated how RDARR compliance for Pillar I models can be operationalised via DCAM in a scalable and evidence-based way for all the credit risk models.

A key finding is that the implementation of an end-to-end data lineage should be treated as the primary architectural lever to accelerate RDARR remediation in the credit risk modelling context as it needs pre-requisites that strengthen governance and data quality controls. Indeed, while IRB models often already include a structured set of data quality controls, the ability to demonstrate auditable traceability from source systems through transformations to model inputs, model outputs, and downstream reporting remains a frequent structural weakness. By designing lineage at the appropriate granularity and explicitly linking it to measurable data quality controls, banks can strengthen not only Principle 4 (Integrated Data Architecture) and Principle 5 (Group-wide Data Quality Management and Standards), but also provide the operational basis for Principle 3 (Effective Data Governance Framework) through clearer ownership boundaries, and enforceable accountability across the end-to-end chain.

Overall, the study contributes to designing a coherent implementation pathway that starts from a high-maturity domain (IRB), uses data lineage as the enabler for linking architecture, quality and governance – effectively reducing the compliance burden that banks often perceive as overwhelming under RDARR principles. Future work should empirically validate the proposed framework through implementation case studies, define standardised RDARR performance indicators for lineage coverage and control effectiveness, and assess its applicability in innovative modelling contexts increasingly shaped by AI-based approaches. These contexts may require enhancing traditional data-quality practices, as consolidated industry and regulatory frameworks may not fully capture new and yet unexplored data-quality dimensions introduced by AI-driven models.

References

Bank for International Settlements (2024). "Digitalisation of Finance: Implications for Governance and Risk". Available at: https://www.bis.org/bcbs/publ/d575.pdf, https://www.bis.org (Accessed: 19 August 2025).

Bank for International Settlements (2026). "Implementation of the Principles for effective risk data aggregation and risk reporting (BCBS 239 Principles)". Available at: https://www.bis.org/publ/bcbs_nl36.html (Accessed: 12 January 2026)

Basel Committee on Banking Supervision (2013). "Principles for effective risk data aggregation and risk reporting (BCBS 239)". Available at: https://www.bis.org/publ/bcbs239.html (Accessed: 19 August 2025).

EDM Council (2014). "DCAM – Data Management Capability Assessment Model". Available at: https://dgpo.org/wp-content/uploads/2016/06/EDMC_DCAM_-_WORKING_DRAFT_VERSION_0.7.pdf (Accessed: 23 November 2025).

EDM Council (2023). "Lloyds Bank achieves competitive advantage with DCAM". Available at: EDMC_Case_Study_Lloyds_Bank.pdf (Accessed: 23 November 2025)

European Banking Authority (2016). "EBA/RTS/2016/03 – Final Draft Regulatory technical standards on Assessment Methodology for IRB". Available at: https://www.eba.europa.eu/activities/single-rulebook/regulatory-activities/credit-risk/regulatory-technical-standards-2 (Accessed: 15 September 2025).

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.thematicreview_riskdataaggregation.en.pdf (Accessed: 19 August 2025).

European Central Bank (2024). "Guide on effective risk data aggregation and risk reporting". Available at: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.supervisory_guides240503_riskreporting.en.pdf (Accessed: 10 June 2025).

European Central Bank (2025). "Sound risk data reporting: key to better decision-making and resilience". Available at: https://www.bankingsupervision.europa.eu/press/supervisory-newsletters/newsletter/2025/html/ssm.nl250219.en.html (Accessed: 10 June 2025).

European Commission (2022). "Commission Delegated Regulation EU 2022/439 of 20 October 2021 supplementing Regulation (EU) No 575/2013 of the European Parliament and of the Council with regard to regulatory technical standards for the specification of the assessment methodology competent authorities are to follow when assessing the compliance of credit institutions and investment firms with the requirements to use the Internal Ratings Based Approach". Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0439 (Accessed: 15 September 2025).

Heß, V. L., & Damásio, B. (2025). "Machine learning in banking risk management: Mapping a decade of evolution". International Journal of Information Management Data Insights, 5(1), 1-17. Article 100324. Available at: https://doi.org/10.1016/j.jjimei.2025.100324 (Accessed: 19 August 2025).

Kobayashi, T. (2016). "Global financial institutions' data governance: implications of EDM Council data management survey". Available at: nri.com (Accessed: 19 August 2025).
McKinsey & Company (2024). "BCBS 239 2.0 resurgence: Strengthening risk management and decision making". Available at: https://www.mckinsey.com/ (Accessed: 1 July 2025).
McKinsey & Company (2025). "Getting the data architecture right in banking". Available at: Getting the data architecture right in banking | McKinsey & Company (Accessed: 19 August 2025).
Number Analytics (2025). "6 Data Quality Strategies Transforming Finance & Banking". Available at: 6 Data Quality Strategies Transforming Finance & Banking (Accessed: 19 August 2025).
Oliver Wyman (2024). "Immediate implications of ECB BCBS 239 Guide. A painful reckoning or a commercial opportunity". Available at: https://coilink.org/20.500.12592/3hgvzzp (Accessed: 1 July 2025).
PwC (2024). "Data Quality Risk Management: Dal presidio dei dati di rischio al presidio del rischio del dato". Available at: https://www.pwc.com/it/it/assets/docs/data-quality-risk-management.pdf (Accessed: 1 July 2025).
PwC (2025). "Il valore espresso da RDARR a beneficio dei programmi di sviluppo della IA", presentation at the ABI conference "Supervision, Risk &Profitability", 10 June 2025 (Accessed: 17 June 2025).
SAS Institute (2024). "Data quality: The Achilles' heel of risk management". Cary, NC: SAS Institute Inc. Available at: https://www.sas.com/en_us/insights/articles/risk-fraud/data-quality-achilles-heel.html (Accessed: 1 July 2025).
Thite, G. (2025). "Modern Data Architectures in Financial Analytics: A Technical Deep Dive". European Journal of Computer Science and Information Technology, 13(22), pp. 79–86. Available at: https://doi.org/10.37745/ejcsit.2013/vol13n227986 (Accessed: 19 August 2025).
University of Oxford (2020). "RACI responsibility matrix: Guidance". Available at https://focus.admin.ox.ac.uk/files/racipdf (Accessed: 07 January 2026).
European Banking Authority (2026). £Final Report – Draft Regulatory Technical Standards on amending Delegated Regulation (EU) No 529/2014 with regard to the assessment of material model changes". Available at: https://www.eba.europa.eu/sites/default/files/2026-03/7bbbbcd4-caa4-489d-b2b4-a837d04e709e/Final%20report%20RTS%20on%20material%20model%20changes.pdf (Accessed: 1 April 2026).

Annex

Here below a summary of RDARR principles stated in the Guide on effective risk data aggregation and risk reporting. European Central Bank 2024:

1. Responsibility of the management body: The ECB stresses that the ultimate accountability for RDARR lies with the management body (board of directors and senior executives). This includes not only approving the RDARR strategy but also ensuring sufficient resources, prioritisation, and integration into the overall risk management framework. Supervisory assessments have shown that insufficient steering and ownership at the top level have been a root cause of slow progress (European Central Bank, 2024). Effective implementation requires clear escalation processes and board-level reporting on RDARR performance indicators.

2. Sufficient scope of application: institutions should establish a data governance framework covering all material entities, risk types—including credit, market, liquidity, operational and third-party risk—and the entire lifecycle of the data. The scope should include internal risk reports used for decision-making and strategic steering, financial reports externally published together with annual financial statements, and supervisory submissions such as FINREP and COREP templates, EBA and SSM stress test exercises, and Pillar 3 disclosures. Furthermore, key risk indicators (KRIs), including risk appetite indicators as well as other key indicators referred to in the internal risk, financial and supervisory reports, should be included in the data governance framework. Lastly, in terms of models, the scope should cover key internal risk management models including, but not limited to, Pillar 1 regulatory capital models (such as internal ratings-based (IRB) approaches for credit risk), Pillar 2 risk and capital models and other key risk management models (such as IFRS9 collective provisions models and value-at-risk models). This includes input data for model development as well as resulting model outputs (e.g. exposure at default, probability of default or loss-given-default estimates) that are crucial for managing the risks faced by the institution.

3. Effective data governance framework: Robust governance is critical for ensuring that risk data are reliable and fit for purpose. According to the ECB Guide, institutions are expected to establish a comprehensive governance framework that operates both at group level and at the level of material legal entities. A key element is the designation of data owners responsible for key risk indicators and critical data elements across the full aggregation process. Their duties include defining and applying data quality controls, ensuring accuracy, integrity, completeness and timeliness, monitoring and reporting data quality, remediating deficiencies, and managing metadata such as data lineage and dictionaries. In addition, a central data governance function should issue policies, oversee implementation across the organisation, and participate in change processes with material impact on RDARR, including mergers, outsourcing, product launches or IT upgrades. The framework must also include a validation function in the second line of defence, independent from RDARR operations, which regularly assesses RDARR processes, IT infrastructure, and outsourced activities, while being adequately resourced and segregated to avoid conflicts of interest. Finally, an internal audit function as the third line of

defence should periodically review the effectiveness of the validation function, the overall governance framework, and the quality of data used for risk quantification, thereby providing independent assurance to management and supervisors.

4. Integrated data architecture: Data architecture and lineage must ensure transparency, traceability, and the minimisation of manual interventions. The ECB Guide (2024) requires institutions to implement and document an integrated data architecture at the group level, supported by harmonised taxonomies and metadata repositories. Supervisory reviews have shown that many banks still rely on fragmented legacy IT systems, resulting in complex reconciliation processes and excessively long production times for risk and financial reports. To address this, supervisory expectations now call for end-to-end documentation of data flows, the adoption of standardised architecture principles, and integrated IT solutions capable of producing consistent datasets across regulatory, supervisory, and internal reporting streams. Such integration also enables automation, reducing the frequency of errors and lowering operational risk. A key element of this framework is data lineage, which refers to the complete documentation of how data move and transform from their initial capture at source systems through intermediate processing (extraction, transformation, and loading) to their final use in risk models, key risk indicators, and reports. Effective lineage makes it possible to verify whether data quality controls are well placed in the flow, to trace back the source of errors or inconsistencies, to ensure that transformations preserve alignment with regulatory definitions, and to provide transparent audit trails for both internal validation and supervisory review. By embedding robust lineage practices, institutions can enhance the reliability, timeliness, and credibility of their risk data aggregation processes.

5. Group-wide data quality management and standards: The ECB requires institutions to establish harmonised data quality (DQ) frameworks across the entire group, embedded within the overall risk management or data governance framework. These frameworks must be supported by adequate data quality controls and measurable data quality metrics covering the key dimensions of accuracy, integrity, completeness, timeliness, and consistency, with systematic monitoring and clear tolerance levels. Institutions are expected not only to detect and report data quality breaches but also to maintain comprehensive registers of deficiencies, conduct root cause and quantitative impact analyses, and implement remediation processes with defined responsibilities, deadlines, and evidence of resolution. The framework must also integrate end-user applications and manual workarounds, ensuring that they are subject to adequate control mechanisms until migrated into IT-controlled environments. In addition, the ECB expects banks to properly consider data quality risks within their ICAAP and ILAAP processes, where unresolved data quality issues might lead to an underestimation of risks and should be reflected in the risk quantification through additional margins of conservatism.

6. Timeliness of internal risk reporting: Risk reports must be delivered frequently and quickly enough to support decision-making during normal and stressed conditions. Supervisory evidence collected by the ECB has shown that, in some institutions, monthly risk reports required over 40 working days to be produced, which is incompatible with effective risk management (European Central Bank, 2018). The ECB expects institutions to ensure that the combination of reporting frequency and production time is calibrated in such a manner as to allow for timely reactions to changes in its risk situation, thereby complying with its set of internal risk appetite indicators (metrics and limits) and ensuring that timeliness does not compromise accuracy or completeness. Moreover, in addition to sound reporting capabilities in normal situations, institutions should implement effective RDARR capabilities for stress or crisis situations to adequately manage unexpected stress events.

7. Effective implementation programmes: The ECB expects banks to manage RDARR remediation through structured, well-resourced programmes with clearly defined milestones, measurable deliverables, and independent oversight, with responsibility for the implementation attributed to the management body. Implementation programmes should include periodic reporting to senior management, external benchmarking, and alignment with broader digitalisation strategies. Good project management practices are essential to achieving tangible progress, as supervisory reviews have shown that many projects lacked sufficient prioritisation, suffered from frequent delays, and failed to integrate lessons from past inspections.