

Dr. Anandkumar Prakasam

Chennai, India

root.anand@gmail.com | +91 97906 36324

15 August 2026

The Chief General Manager

Operational Risk Group, Department of Regulation

Reserve Bank of India, Central Office

Mumbai

Subject: Comments on the draft Guidance on Regulatory Expectations for Data Governance, 2026

Dear Madam or Sir,

I write in response to the invitation for public comments on the draft Guidance on Regulatory Expectations for Data Governance, released for public comments in July 2026. This submission is made in my personal capacity as an independent researcher and practitioner in AI and data governance. It represents no client, employer or commercial mandate. Earlier this month I submitted comments on the draft Guidance on Regulatory Principles for Model Risk Management, and this submission is written as a companion to that one, since the two instruments will govern the same institutions and, in large part, the same systems.

I find the draft well conceived and practical. Its lifecycle-based approach to governance at paragraph 32, its insistence that foundational attributes of data be established at the point of origination at paragraphs 35 and 48, its requirement that disposal be secure, controlled and verifiable at paragraph 42, its extension of lineage through third-party sharing at paragraph 63, and its integration of the Digital Personal Data Protection Act into the governance framework itself at paragraphs 6, 18 and 62, are all sound choices that go further than comparable instruments elsewhere. My comments are offered in that same constructive spirit. Each identifies a place where the Guidance can be made more operational and verifiable without adding compliance burden.

The comments cluster around eight themes. Theme A concerns making the records of data governance demonstrable rather than merely documented. Theme B concerns giving the word verifiable in the disposal clause a working definition. Theme C concerns fixing foundational metadata at origination in a form that cannot be silently rewritten. Theme D concerns reading this Guidance together with the model risk framework and the data protection law, so that one evidence layer serves three instruments. Theme E concerns making deletion of shared data provable across the third-party boundary. Theme F concerns proportionality for the long tail of smaller entities. Theme G concerns reclassification when transformation changes the nature of data, which today has no named owner, no defined trigger and no required record, although nearly every protection in the draft is proportionate to classification. Theme H concerns anchoring the draft's two hardest phrases, intended meaning and fitness for intended use, to the entity's own recorded declarations, so that both become answerable from records.

I would be glad to assist the Reserve Bank further on any of the points raised here, including through the Department of Regulation or any working group that may be constituted as the Guidance is finalised.

Yours faithfully,

Dr. Anandkumar Prakasam

About the Commenter

I hold a PhD in Computer Applications from the National Institute of Technology, Tiruchirappalli, and an LLB (Hons) with a specialisation in regulatory compliance. I have spent twelve years in research and data science consulting, including corporate training engagements for Oracle Financial Services and BHEL, and I have published in Springer journals including Artificial Intelligence Review and Neural Computing and Applications. My current work concerns the governance of data and AI systems at the point where engineering meets law.

I am the author of a five-volume body of work on this subject. The volumes are Decision-Centric AI Governance, on governing the decisions an AI system is permitted to make, Out of Bounds, on normative and white-box red teaming, Human-in-Control, on measurable human oversight, Provenance, on a six-dimensional framework for governing data across its location, its time, its purpose, its influence, its context and its transformations, and The Fiduciary's Burden, a reading of the Digital Personal Data Protection Act, 2023 as an instrument that reverses the burden of proof onto the data fiduciary. The fourth of these volumes bears directly on the subject of this Guidance.

I am also the inventor on a family of seven Indian patent applications relating to verifiable data and AI governance mechanisms, now filed as complete specifications. Those bearing on this Guidance cover tamper-evident evidence generation and forensic reconstruction, lineage-aware deletion with verifiable disposal, cryptographically verifiable decision lineage, continuous authorization with hardware-secured recalibration, and reconciliation of governance obligations issued by more than one authority, and verification that data preserves its declared meaning as it is transformed. Alongside the writing I have built a working reference implementation that demonstrates several of the properties discussed below, in particular tamper-evident records, end-to-end data lineage and verifiable deletion, which I mention only to show that the suggestions here are practical rather than aspirational.

These interests are disclosed for transparency. Nothing in this submission depends on them, and no comment below asks the Reserve Bank to adopt any proprietary method. The properties I refer to are available in commodity form from more than one source, and the Guidance need not name or prefer any technique.

Comments

A. Records of data governance should be demonstrable, not merely documented, paragraphs 15, 19, 33, 41 to 42 and 63

Observation - The draft rightly makes integrity, auditability and traceability first-class expectations at paragraph 15(1), asks for documented ongoing monitoring at paragraph 19, requires lineage from origin through transformation to downstream usage at paragraph 33, and requires that archival preserve linkage to definition, classification and lineage at paragraph 42. These clauses turn on records. A lineage record held in an ordinary mutable repository can be altered, backdated or regenerated after an incident, precisely when its integrity matters most. Supervisory examination would then be comparing the estate's present account of itself with nothing independent of the estate. A decade of experience with risk data aggregation programmes under BCBS 239 suggests that documentation of lineage is achievable while proof of lineage remains rare, and the distance between the two is where examinations stall.

Suggestion - Add, in Chapter II or Chapter IV, an expectation that records material to data governance, at a minimum lineage records, transformation logs, disposal records and the record of data shared with third parties, be maintained in a manner that is **tamper-evident and independently verifiable**, meaning that the entity can demonstrate that a record existed in its present form at a stated past time, without the examiner having to trust the systems that produced it. This raises the standard from we documented the lineage to we can prove what the estate looked like on a given day, which is what paragraph 42's retrieval clause for supervisory and investigative purposes will actually require in practice. The technical means are mature and available from multiple sources, and the Guidance need not prescribe any technique, since the property of verifiable integrity is what matters.

B. The word verifiable in the disposal clause deserves a definition, paragraph 42B

Observation - Paragraph 42B requires that disposal of data be carried out in a secure, controlled and verifiable manner. This is one of the strongest clauses in the draft, and the word verifiable is doing the important work. As drafted, however, it does not say verifiable by whom or from what. An entity could read it as satisfied by an internal disposal log, which is a record of an assertion rather than evidence of an outcome. Deletion is also uniquely hard to evidence, since the natural proof of an action is the artifact it produces, and disposal produces an absence.

Suggestion - Add an explanation that disposal is verifiable when the entity can demonstrate, from records capable of independent verification, that the data was disposed of across every system and copy in which it resided, at a stated time, under a stated authority, with any retained artifact justified against a named legal or regulatory basis. In practice this means the disposal event should generate its own tamper-evident record at the time of execution, including for copies held in backups, archives and derived datasets, so that the completeness of disposal can be examined later rather than asserted. Where a retention obligation prevents disposal of a specific artifact, the basis relied upon should itself be recorded and reviewable, so that what was withheld is as examinable as what was erased.

C. Foundational metadata should be fixed at origination in a form that cannot be silently rewritten, paragraphs 35, 48 and 50

Observation - Paragraphs 35 and 48 are, in my view, the best clauses in the draft. Requiring ownership, classification, usage intent, consent, source system and retention attributes to be established at the point of origination is the single decision that makes everything downstream governable, and paragraph 50's requirement that metadata flow with the data without loss of basic attributes completes it. The remaining gap is temporal. Attributes established at capture but held in ordinary mutable form can be revised later, and after such a revision the entity itself may be unable to establish what the attributes originally were. **The clause that prevents loss of context across systems does not yet prevent loss of context across time.**

Suggestion - Add to paragraph 48 an expectation that the foundational metadata created at capture be preserved in tamper-evident form, so that later corrections and reclassifications append to the record rather than replace it, and the original attributes remain retrievable alongside every subsequent version. This costs little, since the attributes are already being created, and it converts paragraph 55's requirement that classification remain traceable across the lifecycle from an aspiration into a property the entity can demonstrate. It also protects the

entity itself, since the commonest dispute about a data estate is not what the attributes are today but what they were when a decision was taken.

D. Read this Guidance together with the model risk framework and the data protection law, paragraphs 4, 6, 18 and 62

Observation - The draft already reaches across regime boundaries. Paragraph 6(iv) requires the governance framework to comply with the Digital Personal Data Protection Act, 2023 and its Rules, paragraph 18 repeats the requirement for customer data, and paragraph 62(v) carries consent into third-party sharing. At the same time, the institutions in the applicability list are the same institutions that will implement the draft Guidance on Regulatory Principles for Model Risk Management, whose model inventories must record data dependencies and whose documentation must enable traceability and reproducibility. The same credit model's data estate will therefore answer three instruments at once, this Guidance for the estate, the model risk framework for the models the estate feeds, and the data protection law for the personal data within it. Read separately, the three ask for three parallel record-keeping programmes. Read together, they describe one set of facts examined from three directions, and the practical risk for a regulated entity is not the absence of records but contradiction between its own artifacts, a model inventory, a processing register and a lineage record that describe the same system differently to different examiners.

Suggestion - Consider acknowledging this intersection expressly, for example in a note to paragraph 4 on reading the Guidance with other instruments, to the effect that records maintained to demonstrate compliance with this Guidance may serve, and should be reconciled with, the records maintained under the model risk framework and the data protection law. A single evidence layer rendered in three report formats spares the entity three programmes and spares the supervisor the task of adjudicating contradictions between them. Encouraging entities to run a periodic reconciliation across their model inventory, their processing register and their data lineage record, so that the three describe the same systems the same way, would surface the most dangerous class of finding, inconsistency between an entity's own sworn artifacts, before an examiner does.

E. Deletion of shared data should be provable across the third-party boundary, paragraphs 62 and 63

Observation - Paragraph 62 rightly requires systems and controls for the deletion of data shared with third parties, and paragraph 63(i) requires that shared data remain traceable to its designated source with metadata and lineage capturing the extent of sharing. These clauses are ahead of comparable instruments. What they do not yet close is the return path. When an entity instructs a third party to delete shared data, the draft requires the control but not the evidence, so the entity's disposal record for shared data will, in practice, be a copy of its own instruction rather than proof of the outcome. The entity remains responsible under paragraph 60 for data it can no longer see.

Suggestion - Add to paragraph 63 an expectation that deletion of shared data be acknowledged by the receiving party in a form the entity can retain and produce, identifying the data concerned, the systems and copies covered, and the date of execution, with such acknowledgements referenced in the entity's lineage record so that the sharing entry and its closure travel together. Where a third party cannot execute deletion immediately, for example in archival tiers, the pending obligation and its scheduled execution should be recorded rather than left silent. This gives paragraph 60's responsibility a working instrument, aligns the third-

party boundary with the verifiable disposal standard of paragraph 42B, and gives the entity evidence it will need in any event under the data protection law's erasure obligations.

F. Proportionality and pooling for the long tail, paragraph 3

Observation - The applicability list reaches from the largest commercial banks to regional rural banks, urban and rural co-operative banks and base layer NBFCs. The principles are appropriate for all of them. The capacity to stand up a Data Function under an officer of Chief General Manager rank, a board-level committee, an executive committee and the four data roles is not evenly distributed across them, and a small co-operative bank asked for all of it at once is likely to produce a purchased framework document and an unchanged practice. The draft's own proportionality principle at paragraph 15(1)(vi) points in the right direction but is addressed to controls rather than to the implementation burden itself.

Suggestion - Publish an implementation timeline that phases expectations by entity class, with the full organisational structure expected first from the largest entities. Permit smaller entities to meet the Data Function and audit expectations through shared or pooled arrangements, on the model of pooled concurrent audit already familiar in the co-operative sector, subject to independence safeguards. Consider a simplified track for the smallest entities centred on the clauses that matter most and cost least, the foundational metadata at origination, the retention and disposal record, and the third-party sharing record, so that the essential disciplines take root before the full framework is demanded.

G. Reclassification upon transformation needs a named owner, a defined trigger and a record, paragraphs 54 and 55, read with 16, 36, 42, 49, 52, 56 and 62

Observation - Classification is the variable on which nearly every protection in the draft is scaled. Paragraph 16 scales risk assessment to criticality and sensitivity, paragraph 36 scales capture validation to criticality and classification, paragraph 42B scales disposal rigor to criticality, sensitivity and classification, paragraph 49 scales metadata depth to classification, paragraph 52 makes classification the driver of permissible usage and controls, paragraph 56 scales the entire quality regime to it, and paragraph 62 scales third-party controls to it. Paragraph 54 then rightly recognises that aggregation, enrichment and derivation can change materiality, risk profile and sensitivity, and paragraph 55 requires review upon material changes. Read together, these clauses imply that a transformation can change classification and that a classification change re-prices every proportionate control downstream, through to retention and disposal. The draft does not say who is responsible for recomputing classification when this happens, what counts as a material change, or what record of the reclassification must survive. The omission cuts in both directions. A missed escalation leaves sensitive, derived data under-protected across its whole remaining lifecycle. A silent de-escalation invites the argument that derived data is free of its source's obligations, an argument supervisory authority abroad has already rejected in the biometric context.

Suggestion - Four additions would close the gap with machinery the draft already has. First, name the Data Owner, who already holds classification under paragraph 25, as accountable for reclassification upon the events paragraph 54 describes, with conflicts routed through the Data Function's existing role at paragraph 23. Second, define material changes in paragraph 55 to include, at a minimum, transformation, aggregation, enrichment and derivation events, a change in the consuming use of the data, and a change in applicable law. Third, provide that derived data inherits the classification of its most sensitive source until reviewed, so that the default is protective and any downgrade is a recorded decision with reasons. Fourth, require that a reclassification produce a record stating the prior class, the new class, the basis for the change, and the downstream controls affected, and that this record be retained with the data

through retention and disposal, since paragraph 42B's own proportionality depends on the classification being right.

H. Intended meaning and fitness for intended use should be anchored to the entity's recorded declarations, paragraphs 40, 51, 56 and 57

Observation - Paragraph 40(iii) requires that transformation preserve the intended meaning of data and not lead to ambiguity or distortion. Paragraph 57 requires that gaps in data quality not compromise decision-making, risk management, or legal and regulatory reporting. Both expectations are correct, and both, as drafted, ask engineering teams to decide questions that cannot be decided by inspecting the data. Meaning is not intrinsic to a dataset, it depends on the intent under which the data was collected and is used, and whether a quality gap compromises a decision depends on which decision consumes the data. The practical risk is that these two clauses, among the most important in the draft, will be answered by policy assertion, with no record capable of proving either property when a supervisor asks.

Suggestion - Both clauses become answerable from records if the final Guidance anchors them to declarations the draft already requires. For paragraph 40(iii), clarify that the intended meaning of data is the meaning recorded in the entity's own metadata and transformation records, the foundational metadata of paragraph 48 and the purpose and nature of the transformation required by paragraph 51, so that preservation of meaning is assessed against the recorded intent, and the assessment forms part of the impact assessment record that paragraph 40(ii) already requires. For paragraphs 56 and 57, clarify that fitness for intended use is established by quality requirements declared for each consuming decision class and regulatory report, so that compliance with paragraph 57 is evidenced by records showing either that consumption met the declared requirements or that an exception was approved by the accountable owner with reasons. In both cases the judgment remains with the entity's accountable roles, where it belongs, and what the Guidance gains is that the judgment leaves a record an examiner can review.

Closing

The draft Guidance arrives at the right moment and takes the right shape. Its lifecycle spine, its insistence on attributes fixed at origination, its verifiable disposal clause and its extension of lineage through third parties are the choices that will decide whether data governance in regulated entities becomes a practice or remains a binder, and the draft gets each of them right. The eight themes above pursue a single idea. A framework of this kind succeeds when a regulated entity can show, and an examiner can verify, that the estate the records describe is the estate that actually ran. The changes I suggest move the Guidance a short distance further in that direction, from records that are kept to records that can be proved, at little added burden, since in each case the record in question is already being created and the suggestion concerns only the form in which it is held. I hope these comments are useful in taking the Guidance from a strong draft to an enforceable standard, and I remain available to assist.

Respectfully submitted,

Dr. Anandkumar Prakasam
PhD (Computer Applications), LLB (Hons)
Chennai | root.anand@gmail.com | 9790636324