BANK ĊENTRALI TA' MALTA
EUROSISTEMA
CENTRAL BANK OF MALTA

The AI-Powered Foundation for Data Governance to Strengthen Statistics Department

Bilal Kurban[1]

[1] Senior Statistician at the Statistics Department The author would like to thank Taejin Park from BIS for reviewing the paper. The author also would like to thank Jesmond Pule, Karen Caruana and all attendees of the thematic meeting on ESCB/SSM Data Management and Data Governance Network of Experts for feedback and comments provided. The views expressed in this report are those of the author and do not necessarily reflect the views of the Central Bank of Malta.

Abstract

This paper presents an institutional case study of the Central Bank of Malta (CBM) Statistics Department's programme to formalise data governance. The central finding is that strong technical data capabilities do not automatically translate into mature organisational data governance. A structured Data Maturity Assessment (DMA) found comparatively strong performance in systems, data management and data protection, alongside weaker results in knowing the data held, setting data direction and taking responsibility for data. These gaps provide the analytical basis for a four-pillar response: a centralised data inventory and catalogue, maturity assessment, data strategy and governance framework. Artificial intelligence (AI) was used in a deliberately narrow supporting role to generate candidate table descriptions from technical metadata using a locally deployed large language model. In response to the limitations of such metadata-only generation, the revised approach treats the model as a replaceable component, introduces domain grounding and an abstention option for ambiguous inputs, and positions expert validation as a necessary control rather than assuming generated descriptions are authoritative. The paper also proposes measurable implementation indicators that link governance interventions directly to diagnosed maturity gaps. The case suggests that sustainable data governance in statistical organisations requires the combined development of visibility, accountability, strategic direction and proportionate AI controls, while institution-wide scaling depends on cross-functional arrangements and senior organisational sponsorship.

Keywords: data governance; data catalogue; data maturity; metadata; large language models; central banking; official statistics

1. Introduction

Modern statistical organisations operate complex data environments involving multiple sources, processing systems, transformation stages and organisational responsibilities. Strong infrastructure and validation processes are essential, but they do not by themselves ensure that data assets are easy to find, consistently documented, clearly owned or governed according to a common strategic direction. This distinction between technical capability and organisational governance provides the central motivation for the initiative examined in this paper.

In July 2025, following discussions at the CBM's 46th Statistics Committee meeting, the Statistics Department began formalising its data governance practices. The department did not start from an absence of controls. Established statistical production processes, technical infrastructure, validation mechanisms and data-protection practices were already in place. The challenge was that important governance practices remained distributed across systems, units and individual expertise rather than integrated within a common, visible and accountable structure.

The subsequent DMA provided a structured way to quantify these concerns. It showed a department with advanced operational capabilities but weaker formal governance in three areas in particular: findability and knowledge of internal data holdings, strategic alignment, and responsibility for data. The initiative is therefore framed in this paper not primarily as a compliance response, but as an institutional response to diagnosed governance gaps. European and national frameworks remain important requirements and reference points that the resulting governance model must satisfy.

1.1 External Alignment and Institutional Constraints

The CBM operates within the Eurosystem statistical environment and aligns its statistical work with the European Central Bank's Statistical Quality Framework. Preparations for increasingly integrated reporting arrangements also increase the importance of harmonised metadata, repeatable processing and clear documentation. At national level, Malta's Public Administration Data Strategy 2023–2027 emphasises transparency, data literacy, interoperability, accountability and responsible reuse. International central-bank discussions similarly point to growing interest in formal data strategies, catalogues and governance as foundations for wider use of Big Data and AI.

These external frameworks are treated here as design constraints and alignment mechanisms rather than substitutes for institution-specific diagnosis. Effective governance must respond to the actual data flows, responsibilities and weaknesses of the organisation in which it is implemented.

1.2 Analytical Proposition and Four-Pillar Response

The case is organised around a simple proposition: technical data maturity and organisational data-governance maturity are related but distinct. The programme therefore links each diagnosed need to a corresponding intervention. Visibility and findability are addressed through the inventory and catalogue; measurement is provided by the DMA; strategic direction is addressed through the Data Strategy; and accountability is addressed through the Governance Framework. The four pillars are mutually reinforcing rather than independent projects.

Diagnosed need	Governance response	Purpose
Limited visibility/findability	Inventory & Data Catalogue Explorer	Create a searchable view of statistical assets, metadata and lineage
Need for objective baseline	Data Maturity Assessment	Identify strengths, gaps and priorities using a structured framework
Weak strategic alignment	Statistics Department Data Strategy	Translate diagnosis into priorities, sequencing and measurable actions
Unclear/formal responsibility	Governance Framework	Clarify ownership, stewardship, custody and escalation

2. Methodology and Implementation

2.1 Pillar I: Data Inventory and Catalogue

The first workstream created a centralised inventory intended to serve as a single point of reference for statistical data assets. The inventory covers 203 surveys and more than 7,000 tables and brings together information on surveys, tables, columns and dependencies. A custom Data Catalogue Explorer was developed in Python and Streamlit to make this metadata searchable and usable by staff.

The application uses pandas for data handling, Streamlit for the interface and PyVis for interactive lineage visualisation. It provides three principal navigation levels: a Survey Explorer, a Table Explorer and a Column Explorer. Parent-child relationships between datasets can be displayed as interactive lineage graphs, allowing users to explore how tables relate to upstream and downstream assets. A sidebar reports inventory statistics such as numbers of surveys, tables, columns and dependencies.

These are demonstrable implementation outputs. Wider benefits such as reduced search time, lower dependence on individual knowledge and improved collaboration are plausible objectives of the catalogue but were not quantitatively measured during the initial implementation; they are therefore treated as outcomes for subsequent evaluation rather than as demonstrated impacts.

2.1.1 AI-Assisted Candidate Metadata Generation

A practical challenge in constructing the inventory was that some database tables lacked sufficiently informative business descriptions. A locally deployed large language model was therefore used to generate concise candidate descriptions from table names and column names. The objective was not to infer authoritative business meaning autonomously, but to accelerate first-pass metadata documentation and identify records requiring expert attention.

The prototype implementation used Ollama to execute Llama 3 8B locally. Local execution was selected because it allowed metadata to remain within the Bank’s controlled environment. The choice of Llama 3 8B should be understood as a practical prototype choice rather than evidence that it is the optimal model for the task. The methodological design is better treated as model-agnostic: metadata are prepared, passed through a prompt-construction layer and a replaceable model interface, after which the candidate output is subject to validation before entering the catalogue.

This separation is important for maintainability. Stronger locally deployable models can be substituted without changing the surrounding catalogue workflow, and future model comparisons can use identical inputs and evaluation criteria.

2.1.2 Prompt Design, Ambiguity and Abstention

The initial prototype prompt[2] asked the model to infer a one-sentence description from the table and column names. Peer review highlighted two important limitations: the prompt provided little domain context, and it forced a description even when names were opaque. A metadata-only model cannot reliably infer business meaning from fields such as FLAG1, FLAG2 or generic identifiers when the same information would also be unclear to a human reviewer.

The revised prompt design[3] therefore grounds the task explicitly in central-bank statistical data, provides examples of acceptable descriptions, instructs the model not to invent unsupported information and introduces an abstention response — "Unclear — requires manual review" — when the available metadata are insufficient. This changes the role of the LLM from an automatic description generator to an AI-assisted metadata triage mechanism: sufficiently informative inputs can receive a candidate description, while ambiguous inputs are escalated for human review.

A further methodological step is to assess the informativeness of the source metadata itself. A representative sample of table and column names can be classified as self-explanatory, partially informative or opaque before evaluating model output. This makes it possible to distinguish model error from a more fundamental limitation in the metadata supplied to the model.

2.1.3 Validation and Evaluation of AI-Generated Metadata

To evaluate the reliability of the AI-generated table descriptions, a systematic review was conducted across all 2,920 generated descriptions included in the evaluation. Each description was assessed against the available information about the corresponding table and classified into four categories: (i) **correct**, where the description accurately represented the table without requiring substantive modification; (ii) **partially correct**, where the general meaning was captured but clarification or substantive editing was required; (iii) **incorrect**, where the description materially misinterpreted the table or introduced unsupported information; and (iv) **insufficient**

---

[2] "Given a table named '{table_name}' with the following columns: '{column_names}'. Provide a one-sentence, concise description of the data this table might contain. Be helpful and professional, do not use bullet points or lists."

[3] You are assisting with metadata documentation for statistical data tables used within a central bank.
Your task is to generate a concise, one-sentence description of the likely content of a table using only the table name and column names provided.
Do not invent information that cannot reasonably be inferred from the metadata. If the table name and column names do not provide sufficient information to determine the table's likely purpose, return exactly:
"Unclear — requires manual review."
Examples:
Table: EXCHANGE_RATES
Columns: OBS_DATE, CURRENCY, EXCHANGE_RATE
Description: Contains exchange-rate observations by reference date and currency.
Table: SECURITIES_HOLDINGS
Columns: OBS_DATE, ISIN, COUNTERPART_AREA, MARKET_VALUE
Description: Contains statistical information on securities holdings by reference date, security identifier, counterpart area and market value.
Table: TAB_X17
Columns: ID, FLAG1, FLAG2, VAL
Description: Unclear — requires manual review.
Now describe the following table:
Table: {table_name}
Columns: {column_names}
Return only the description and no additional explanation.

metadata, where the available table and column names did not provide enough information to determine the table's meaning reliably.

Assessment	Original Prompt	Revised prompt
Correct	16%	13%
Partially correct	37%	45%
Incorrect	46%	33%
Insufficient metadata	1%	8%
Total	**100%**	**100%**

The results show that the revised prompt did not increase the proportion of descriptions classified as fully correct, which decreased slightly from 16% to 13%. Its main benefit was instead a reduction in clearly incorrect descriptions, from 46% to 33%. At the same time, partially correct descriptions increased from 37% to 45%, while cases identified as having insufficient metadata increased from 1% to 8%.

This pattern is important from a data-governance perspective. The objective of the revised prompt was not simply to maximise the number of automatically accepted descriptions, but also to reduce unsupported inference and make uncertainty more visible. The increase in partially correct and insufficient-metadata classifications, together with the reduction in incorrect descriptions, suggests that the revised approach is more conservative in situations where the available technical metadata do not support a confident interpretation.

Nevertheless, the results also demonstrate that AI-generated descriptions cannot currently be treated as authoritative metadata. Only 13% of the descriptions produced using the revised approach were judged fully correct without substantive modification, while 33% remained incorrect. Human validation therefore remains an essential component of the workflow. In operational use, the model is best positioned as an assistant for generating candidate metadata, with subject-matter experts retaining responsibility for validation, correction and approval.

The evaluation also highlights a limitation that is not attributable solely to model performance. Because the model receives only table and column names, ambiguous abbreviations, generic field names and system-generated identifiers may provide insufficient semantic information for either the model or a human reviewer to determine the table's business meaning reliably. Future development should therefore consider enriching the model input with additional contextual metadata, such as survey descriptions, existing business definitions, related tables and lineage information, rather than relying exclusively on increasingly complex prompting.

2.2 Pillar II: Data Maturity Assessment

The Statistics Department applied the UK Government Data Maturity Assessment framework to establish a structured baseline. The framework covers ten topic areas across six broader themes and uses five maturity levels: Beginning, Emerging, Learning, Developing and Mastering. The assessment was conducted on 28 November 2025 in a collaborative workshop in which department managers scored the assessment items through collective discussion and voting.

The overall average was 3.9. The results show an important asymmetry: technical and protective capabilities were relatively strong, while several organisational governance dimensions were weaker.

Topic	Average score	Interpretation in assessment
Having the right systems	4.3	Developing – strength
Managing your data	4.2	Developing – strength
Protecting your data	4.2	Developing – strength
Engaging with others	4.1	Developing
Managing and using data ethically	4.0	Developing

Having the right data skills and knowledge	3.8	Developing
Making decisions with data	3.7	Developing
Knowing the data you have	3.5	Learning – priority
Setting your data direction	3.5	Learning – priority
Taking responsibility for data	3.1	Learning – priority

At a more granular level, sharing data with stakeholders scored 4.8, while creating and embedding data governance scored 2.4. The results therefore do not describe an organisation lacking data capability. Instead, they indicate that technical systems and operational practices have developed further than formal ownership, strategic alignment and institutionalised governance. This diagnosis provides the rationale for the remaining pillars.

2.3 Pillars III and IV: Data Strategy and Governance Framework

The DMA findings are being used to shape the Statistics Department Data Strategy and the Governance Framework. The Data Strategy translates the diagnosis into priorities around data quality, regulatory alignment, transparency, accountability, metadata and accessibility. The Governance Framework is intended to formalise responsibility for statistical data and establish mechanisms for stewardship, ownership, technical custody and escalation.

Three operational roles are distinguished. The Data Owner is accountable for the meaning, content and quality of a data asset; the Data Steward supports its day-to-day governance and metadata; and the Data Custodian is responsible for the relevant technical environment. Defining these roles is different from assigning them in practice. Within the Statistics Department, assignment can be mapped more directly to statistical domains and production processes. For assets that cross departmental boundaries, ownership and stewardship can be more complex and require coordination with IT and other business areas.

Accordingly, the framework should initially be understood as a departmental governance model and a foundation for possible wider institutional adoption, not as evidence that bank-wide ownership questions have already been resolved.

3. From Maturity Gaps to Measurable Outcomes

The original project narrative described a number of expected benefits, but an academic evaluation requires a clearer distinction between outputs already delivered, measurable implementation indicators and longer-term outcomes. The revised approach therefore links the main DMA gaps to specific interventions and indicators.

DMA finding / need	Response	Interim indicator	Longer-term evidence
Creating and embedding governance (2.4)	Governance Framework	Framework milestones completed; governance bodies/processes established	Improvement in corresponding DMA item
Taking responsibility for data (3.1)	Owner/Steward/Custodian model	% of priority assets/domains with formally assigned roles	Higher responsibility maturity and fewer unresolved ownership issues
Knowing the data you have (3.5)	Inventory & Catalogue	% of identified assets catalogued; % with validated descriptions; lineage coverage	Improved findability and next DMA result
Setting data direction (3.5)	Data Strategy & roadmap	% of agreed strategic actions completed	Improved strategic-direction maturity

Metadata documentation	AI-assisted candidate descriptions	Descriptions generated; accepted/edited/rejected/abstained after validation	Measured metadata quality and catalogue usefulness
Existing strengths (4.0–4.3)	Preserve effective controls	Continued operation of systems, protection and management controls	Maintenance of strength in future DMA

3.1 Short-Term Outcomes and Interim Communication

Short-term monitoring is important because changes in organisational maturity take time. Tangible implementation outputs can demonstrate progress and maintain engagement while longer-term effects are still developing. The initial programme has already established a catalogue covering 203 surveys and more than 7,000 tables, a searchable Explorer, AI-assisted candidate descriptions and interactive lineage functionality.

The next stage should quantify catalogue coverage, validated-description coverage, lineage coverage, assignment of governance roles and completion of roadmap actions. The AI-validation exercise described above should report accepted, edited, rejected and abstained descriptions rather than relying on a general statement that the model worked well. Periodic communication of these indicators to management and staff can provide evidence of progress and help sustain organisational momentum.

Benefits such as reduced search time, improved collaboration and reduced duplication should be evaluated separately using appropriate baselines or user measures before being reported as achieved impacts.

3.2 Medium- and Long-Term Outcomes

At medium term, the programme aims to strengthen structured dissemination, cross-institutional collaboration and consistency of data management. At longer term, the vision is a controlled Statistical Data Space in which datasets, metadata, lineage, quality controls and governance responsibilities are connected within a coherent environment. These remain strategic objectives rather than demonstrated outcomes of the initial implementation.

A subsequent DMA provides a natural mechanism for testing whether weaker dimensions have improved while existing strengths have been maintained. Reassessment is therefore an important part of the governance cycle rather than a one-off maturity exercise.

4. Challenges and Lessons Learned

4.1 Breaking Organisational Silos Requires Governance Beyond Documentation

A centralised inventory and catalogue improve visibility but do not by themselves eliminate organisational silos. Statistical data flows frequently cross functional boundaries involving data producers, statistical compilers, IT infrastructure and downstream users. Effective governance therefore requires mechanisms that operate across those boundaries.

At departmental level, common metadata standards, documented lineage and defined ownership and stewardship can improve coordination. Shared assets require additional arrangements: cross-functional governance structures, common metadata and quality standards, explicit accountability, and mechanisms for escalating and resolving cross-departmental issues. These changes go beyond strengthening documentation.

This distinction matters for scaling. The Statistics Department can develop, test and demonstrate the governance model within its remit, but institution-wide adoption requires organisational commitment and senior management sponsorship capable of establishing common responsibilities and standards across departmental boundaries. Senior sponsorship is therefore treated as a prerequisite for wider adoption, not as an outcome already demonstrated by the departmental project.

4.2 AI Is Useful When Its Role Is Narrow and Controlled

The metadata use case illustrates where AI can add practical value without replacing subject-matter accountability. LLMs can accelerate first-pass documentation, but the quality of the result depends on the informativeness of the input metadata, prompt design and model capability. The model should therefore be replaceable, uncertainty should be explicit, and human validation should remain part of the workflow.

More complex agentic or retrieval-augmented workflows could enrich future descriptions with survey definitions, existing metadata and lineage context. Such approaches may improve performance, but they also introduce additional complexity, evaluation requirements and governance considerations. They are therefore treated as potential extensions rather than prerequisites for the initial catalogue.

4.3 Formalise the Informal, but Do Not Confuse Documentation with Governance

The initiative found that several strong practices already existed but were embedded in operational routines. Formal documentation can make these practices more transparent, repeatable and resilient to staff changes. However, documentation alone does not establish accountability. Effective governance also requires assigned decision rights, escalation paths, monitoring and organisational sponsorship.

4.4 AI Controls Should Be Proportionate to the Use Case

The current AI application is deliberately narrow: it generates candidate descriptions of statistical tables from technical metadata. It does not produce official statistics, make data-quality decisions or support individual-level or policy decisions. The principal risks are therefore inaccurate or unsupported semantic interpretation, insufficient source metadata, confidentiality, traceability and over-reliance on generated text.

Proportionate controls include local processing, explicit labelling of generated text as candidate metadata, abstention when evidence is insufficient, expert validation and monitoring of validation outcomes. A comprehensive institutional AI governance framework is beyond the scope of this implementation. Such a framework would become increasingly important if AI were extended to higher-impact activities such as automated statistical production, analytical interpretation or policy-support processes.

5. Discussion

The CBM case suggests that the most important governance challenge was not a lack of technical capability but a mismatch between operational maturity and formal organisational governance. This matters for statistical institutions because investment in infrastructure, automation and validation can create an impression of overall data maturity even when ownership, discoverability and strategic direction remain underdeveloped.

The four-pillar approach provides a practical sequence for addressing that mismatch. The inventory makes the data environment visible; the DMA converts perceptions into a structured diagnosis; the strategy translates diagnosis into priorities; and the governance framework assigns responsibility for implementation. AI can support this sequence, but its contribution should be evaluated as a component of the metadata workflow rather than treated as evidence of governance maturity in itself.

The case also highlights the importance of scope. Departmental governance can be implemented through local roles, standards and tools, whereas institution-wide governance requires authority across organisational boundaries. This is why senior sponsorship and cross-functional mechanisms are central to scaling the model.

There are limitations. The DMA reflects a structured internal assessment rather than an external audit. The initial AI metadata exercise did not include a formal quantitative benchmark, and user-level benefits such as reduced search time were not measured. These limitations constrain claims about impact but also define a clear evaluation agenda: validate AI-generated metadata, establish catalogue and governance KPIs, collect user evidence where appropriate, and repeat the DMA to measure change over time.

6. Conclusion

The Central Bank of Malta Statistics Department's experience shows that a technically capable statistical organisation can still face important governance gaps. The DMA identified a pattern of relatively strong systems, data management and protection alongside weaker findability, strategic direction and responsibility for data. Reframing the initiative around this diagnosis creates a clearer rationale for the four-pillar programme.

The centralised inventory and Data Catalogue Explorer provide the visibility layer; the DMA provides the baseline; the Data Strategy provides direction; and the Governance Framework provides the basis for accountability. AI contributes in a supporting role by generating candidate metadata at scale, but reliable use requires domain-aware prompting, an option to abstain, model replaceability and expert validation.

The next phase should focus less on broad claims of impact and more on measurable implementation evidence: catalogue coverage, validated metadata, lineage coverage, role assignment, roadmap completion and subsequent maturity reassessment. At the same time, wider institutional adoption will require cross-functional governance and senior management sponsorship beyond the Statistics Department.

The broader lesson for statistical organisations is that data governance is not created by technology, documentation or compliance alone. Sustainable governance emerges when technical capability is connected to discoverability, explicit accountability, strategic direction, measurable outcomes and proportionate controls for emerging technologies.

References

Banque de France / International Banking and Finance Institute (IBFI). Survey on Big Data and Artificial Intelligence in Central Banks, May 2025.

Central Bank of Malta. Statistics Department Data Maturity Assessment results and internal governance project documentation, 2025–2026.

European Central Bank. Statistical Quality Framework.

Government of Malta. Public Administration Data Strategy 2023–2027.

UK Government. Data Maturity Assessment Framework.