

From Data Defects to Decision Failure: A Practical Framework for Data Quality Management in Emerging Markets

Olajumoke Ore Akinremi

Email: akinremiolajumoke77@gmail.com

Abstract

Data quality is a persistent constraint on the reliability of information systems and analytics, yet much of the dominant guidance on data quality management implicitly assumes stable infrastructure, high automation, and mature governance capacity. In many emerging market organisations, data is produced and maintained through hybrid arrangements that combine manual capture, spreadsheet mediation, partial system integration, and locally improvised routines that compensate for gaps in process design and institutional capacity. In these settings, data quality failure is not only a technical defect in stored records. It functions as a decision risk pathway through which duplication, incompleteness, ambiguity, inconsistency, and delayed updates propagate into operational error, managerial misjudgment, and accountability breakdown. This paper develops a coherent, publication-oriented first draft that links established data quality theory and governance principles to the institutional realities of resource-constrained implementation environments. The manuscript is grounded in consumer-oriented data quality theory (Wang & Strong, 1996; Strong et al., 1997), the ontological anchoring of intrinsic data quality (Wand & Wang, 1996), data governance and stewardship principles as articulated in DAMA DMBOK (DAMA International, 2017), and the ICT4D lens on design actuality gaps and local improvisation (Heeks, 2002). Methodologically, the paper adopts a design science research stance (Hevner et al., 2004; Peffers et al., 2007) and treats PRO DQ as the central artifact. PRO DQ is a circular framework comprising Prevent, Responsibilise, Observe, Diagnose, and Fix. The framework is designed to be implementable under constrained budgets and mixed digital maturity, while remaining auditable and governance aware. The paper contributes a defect taxonomy, a defect-to-decision propagation model, and a practical implementation logic that can be evaluated in future field studies through case-based deployment, pilot interventions, and longitudinal monitoring.

Keywords: Data Quality; Data Governance; Design Science Research; ICT4D; Decision Risk; Emerging Markets; Stewardship; Information Systems.

How to Cite: Akinremi, O. O. (2026). From Data Defects to Decision Failure: A Practical Framework for Data Quality Management in Emerging Markets. *Advanced Journal of Science, Technology and Engineering*, 6(2), 13–26. <https://doi.org/10.52589/AJSTE-PI1F7GMV>

INTRODUCTION

Data has become an operational and institutional resource. Organisations use data to coordinate frontline service delivery, manage inventories and logistics, administer benefits and housing, allocate budgets, and evaluate performance. At the same time, public policy debates increasingly treat data as a strategic input into national development agendas and service reform. These shifts are visible across many emerging markets, including African contexts. Still, the practical meaning of “data-driven” decision-making is shaped by how data is produced and governed on the ground. When records are incomplete, duplicated, inconsistent, or updated too late to inform action, data does not simply become less accurate. It becomes less governable, less trustworthy, and less useful for decision-making.

A defining feature of many low- and middle-income organisational settings is the persistence of hybrid information infrastructures. Data capture may still begin on paper, then be transcribed into a system, and later copied into spreadsheets for reporting and reconciliation. Even where digital systems exist, they may be fragmented across units, implemented at different times, and connected through manual exports rather than robust integration. Frontline workers often operate under throughput pressure and limited training time, which increases the likelihood of inconsistent entry and the use of local workarounds. Spreadsheet dependence can be rational from a local perspective because it enables speed and flexibility, but it often weakens standardisation, version control, and traceability. These conditions create a sociotechnical context in which data quality is continuously produced and degraded through routine work.

The mainstream data quality literature provides a strong foundation for understanding what quality is and why it matters. The influential “fitness for use” formulation emphasises that data quality is judged by consumers relative to a task context (Wang & Strong, 1996). This is especially relevant in emerging markets where data is often expected to support high-stakes operational decisions under uncertainty. The ontological view of intrinsic data quality further argues that some defects can be understood as failures of representation between the real world and its encoded form (Wand & Wang, 1996). These perspectives clarify why the same dataset may be considered adequate for one purpose and inadequate for another, and why defects such as ambiguity and incompleteness can be particularly harmful when a decision requires clear classification or timely intervention.

At the managerial level, established governance frameworks emphasise that sustained quality depends on roles, controls, monitoring, and lifecycle management rather than one-time cleaning (DAMA International, 2017). Yet many data quality management models implicitly presume stable infrastructure, reliable integration, consistent enforcement, and specialised teams. In many emerging market organisations, these assumptions do not hold. Governance roles may exist in name but lack authority, resourcing, or clarity. Controls may exist informally but not be documented or measured. Monitoring may be sporadic because producing the monitoring report competes with operational workload. When quality management designs do not match actual work conditions, they are difficult to implement and maintain.

This mismatch is well captured by ICT4D research on design actuality gaps. Heeks (2002) argues that information systems in developing contexts often fail, or underperform, when imported designs are misaligned with local realities in infrastructure, skills, incentives, and institutional capacity. In response, local improvisation emerges as a practical necessity. Improvisation can keep services running, but it can also introduce hidden dependencies and undocumented process variation that destabilises data definitions and traceability. In such settings, data quality failure should be conceptualised as a decision risk pathway. Defects

introduced during capture, transfer, reconciliation, or interpretation can propagate across systems and routines until they culminate in a decision error, an accountability dispute, or an operational failure.

This paper addresses a gap in the literature and practice. While theory explains data quality dimensions and governance frameworks explain managerial responsibilities, there is limited integrative guidance that explicitly models how defects propagate into decision failure in resource-constrained settings and how interventions can be sequenced for implementability. The paper's objective is to develop a theoretically grounded and implementable framework for data quality management in emerging markets that links defect management to decision reliability.

The guiding research question is:

How can organisations operating under resource constraints manage common data defects in ways that reduce decision failure risk and strengthen governance capacity?

The paper contributes a conceptual and practice-oriented artifact called PRO DQ. PRO DQ is designed to support organisations with constrained budgets, mixed digital maturity, and heavy human mediation of data flows. The remainder of the manuscript reviews relevant theory, explains the design science stance, presents the defect taxonomy and propagation model, and develops the PRO DQ framework along with implementation guidance and metrics.

LITERATURE REVIEW

Data Quality Theory and “Fitness For Use”

Data quality research has repeatedly emphasised that quality is multidimensional and context-dependent. Wang and Strong (1996) identify intrinsic, contextual, representational, and accessibility dimensions that data consumers use to judge quality. This framing remains influential because it shifts attention away from a narrow focus on accuracy and towards the practical question of whether data supports a specific task. In operational environments, completeness and timeliness often matter as much as accuracy because decisions occur within time windows and depend on sufficient contextual detail. Strong et al. (1997) further reinforce that quality judgments arise in context, which is particularly relevant when different organisational units interpret the same metric through different operational logics.

A second important anchor is the ontological approach to intrinsic data quality. Wand and Wang (1996) argue that intrinsic defects can be understood as representational failures such as incompleteness, ambiguity, and incorrectness relative to the real-world phenomena being modeled. This view is valuable in emerging market settings where the representation chain is often long and human-mediated. A record may be “incorrect” not because the database engine malfunctioned, but because frontline observation was ambiguous, transcription was rushed, identifiers were missing, or local codes were used inconsistently. The ontological lens also highlights why some defect categories, particularly ambiguity and incompleteness, are not merely nuisances but structural threats to interpretability and subsequent action.

Governance and Stewardship Perspectives

If data quality is fitness for use, then sustainable quality management requires organisational arrangements that make fitness for use an ongoing responsibility. DAMA DMBOK provides

a comprehensive reference model for data management, including governance, stewardship, metadata, quality, and controls (DAMA International, 2017). The key insight relevant here is that quality is a lifecycle property produced through policies, standards, defined roles, escalation routes, and monitoring routines. Khatri and Brown (2010) argue that governance is fundamentally about decision rights and accountabilities. This is crucial in environments where quality issues are persistent because without clear ownership, defects are treated as unfortunate events rather than governable problems.

Governance in practice also involves monitoring and issue management. Pipino et al. (2002) describe approaches to assessing data quality that include both objective and subjective measures. Their emphasis on systematic assessment supports the idea that quality improvements require measurement. However, in many resource-constrained settings, measurement itself is costly because data is fragmented and reporting tools are limited. This increases the need for lightweight but meaningful metrics that are aligned to decision use cases rather than to an exhaustive catalog of quality dimensions.

Data Defects, Propagation, and Organisational Consequences

The data quality and data cleaning literature provides methods for detecting and correcting defects such as duplication, inconsistency, and invalid values. Rahm and Do (2000) review data cleaning approaches and clarify that cleaning is both technically complex and organisationally consequential because it changes the meaning and trustworthiness of downstream information. Batini and Scannapieco (2006) similarly show that data quality problems are multifaceted and require integrated approaches that combine technology and management. Redman (1998) argues that poor data quality has broad organisational costs and that defects tend to propagate and replicate across systems and processes. The relevance of this argument for emerging markets lies in the way hybrid infrastructures and spreadsheet replication can amplify propagation. Once a defective record is exported to spreadsheets, emailed across teams, and copied into new files, correction becomes difficult, and accountability becomes diffuse.

ICT4D and the Design Actuality Gap

Information systems research in developing contexts provides a strong lens for understanding why standard models of data management and governance may underperform in emerging markets. Heeks (2002) conceptualises the design actuality gap as the difference between assumptions embedded in an information system design and the realities of local implementation, including infrastructure, skill levels, processes, and institutional norms. Local improvisation can narrow the gap, but it can also entrench informal routines and weaken standardisation. For data quality management, this implies that defect pathways are often created through reasonable local adaptations, such as keeping parallel spreadsheets to ensure continuity during system downtime, or bypassing validation rules to reduce queue times. These adaptations are not simply user error. They are responses to constraints, but they can create systemic risks when decisions depend on stable definitions, timeliness, and traceability.

Synthesis and Conceptual Gap

Across these literatures, several points converge. Data quality is multidimensional and context-dependent (Wang & Strong, 1996). Intrinsic quality can be understood as representational fidelity and interpretability (Wand & Wang, 1996). Governance requires explicit decision rights, stewardship, and monitoring (DAMA International, 2017; Khatri & Brown, 2010). In

developing contexts, imported designs can fail when they ignore local realities and improvisation (Heeks, 2002). What is less developed is an integrative framework that makes the defect-to-decision pathway explicit and provides an implementable sequence of interventions suited to constrained environments. Existing models often describe what quality is and what governance should look like. Still, they less often provide a coherent logic that links defect types, propagation mechanisms, decision failures, and a feasible cycle of prevention, accountability, monitoring, diagnosis, and correction. The next chapters develop this missing link through a design science artifact called PRO DQ.

METHODOLOGY

Design Science Research

This paper adopts a design science research (DSR) stance. DSR is appropriate when the research goal is to create a purposeful artifact that addresses a relevant organisational problem while being grounded in established knowledge (Hevner et al., 2004). In the present study, the organisational problem is persistent data defects that undermine decision reliability in emerging market organisations operating under constrained resources and mixed digital maturity. The artifact is the PRO DQ framework, supported by a defect taxonomy and a defect-to-decision propagation model.

Design science emphasises both relevance and rigor. Relevance is established by the practical importance of data for organisational coordination and public service delivery, and by the observed mismatch between the assumed conditions of many data quality programs and the realities of hybrid information infrastructures. Rigor is achieved by anchoring the artifact in established theory and frameworks. The paper draws on consumer-oriented data quality theory (Wang & Strong, 1996; Strong et al., 1997), ontological foundations of intrinsic quality (Wand & Wang, 1996), governance and stewardship principles (DAMA International, 2017; Khatri & Brown, 2010), and ICT4D perspectives on design actuality gaps and improvisation (Heeks, 2002). The artifact is designed to be conceptually coherent, implementable, and auditable, meaning that organisations can not only apply it but also produce evidence of application through documented roles, metrics, and issue logs.

The design process aligns with common DSR methodologies that move from problem identification to objectives, design and development, and communication (Peppers et al., 2007). The problem is specified as a decision risk pathway originating in data defects and amplified by fragmented systems, weak controls, unclear stewardship, and local workarounds. The objectives are to reduce defect generation, improve accountability, increase visibility into quality in motion, support diagnosis of root causes, and enable correction that prevents recurrence. The artifact is designed as a circular framework rather than a linear checklist, reflecting the reality that quality is not a one-time achievement but a sustained organisational capability.

This manuscript is conceptual design research rather than empirical validation research. It does not report new quantitative findings, case outcomes, or measured effect sizes. Claims about organisational outcomes are therefore framed conservatively and analytically, focusing on plausible mechanisms supported by the literature. Future evaluation of PRO DQ can proceed through multiple strategies. First, qualitative case studies can examine whether PRO DQ improves role clarity and reduces reliance on informal workarounds, and whether defect propagation becomes more traceable. Second, pilot deployments can monitor changes in

targeted defect metrics such as duplication rates or completeness of critical fields, and can assess whether decision cycle times change for specific processes. Third, comparative field testing across sectors can examine how contextual factors such as regulatory oversight, staffing stability, and integration maturity influence implementation. These evaluation strategies are consistent with DSR's emphasis on utility and contextual performance (Hevner et al., 2004).

THE PRO DQ FRAMEWORK

PRO DQ is a circular framework designed to manage data quality as a decision reliability capability in resource-constrained environments. The framework has five phases: Prevent, Responsibilise, Observe, Diagnose, and Fix. The phases are sequenced in a way that reflects an implementation reality in which prevention and responsibility are prerequisites for effective monitoring, and monitoring is a prerequisite for diagnosis and sustainable correction. The cycle is circular because quality work generates learning that should feed back into better prevention and clearer responsibility over time.

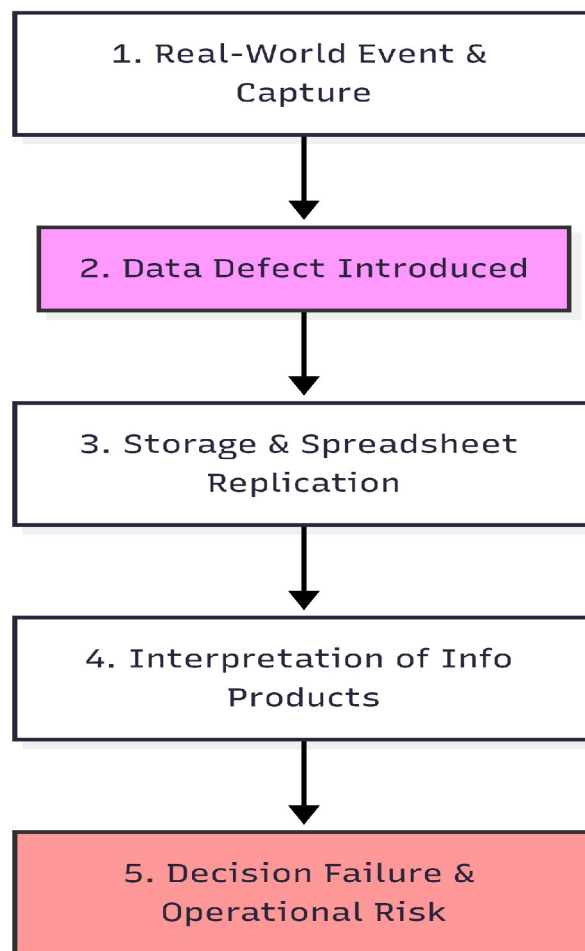


Figure 1: Defect to Decision Propagation Flow

Figure 1 makes explicit the claim that defects become consequential through propagation. In hybrid infrastructures, the defect may not be visible at capture, but it becomes operationally salient once it shapes a queue, a payment decision, a prioritisation list, or a performance metric.

Table 1: Taxonomy of Six Data Defect Categories in Emerging Markets

Defect Category	Definition in Practice	Typical Organisational Conditions	Decision Risk Mechanism
Duplication	Multiple records represent the same entity	weak identifiers, re-entry during outages, parallel registers	inflated counts, double servicing, incorrect segmentation
Incompleteness	Required fields or contextual attributes are missing	Paper-to-system transcription, rushed intake, unclear mandatory rules	inability to assess risk, misprioritisation, manual rework
Ambiguity	Labels, codes, or definitions permit multiple interpretations	weak metadata, inconsistent code lists, informal categories	misclassification, inconsistent execution across teams
Inconsistency	Conflicting values across systems or within a dataset	fragmented applications, spreadsheet copying, ungoverned edits	contradictory reports, loss of trust, decision delays
Delay (timeliness)	Updates occur too late for the decision window	batch updates, manual reconciliation, slow approvals	late interventions, stale dashboards, reactive management
Workaround drift	local improvisations become stable but undocumented routines	design actuality gaps, weak enforcement, staff turnover	hidden process variation, weak auditability, unstable KPIs

The taxonomy is deliberately compact. Its purpose is to identify defect categories that frequently occur when human mediation and fragmentation are persistent. It also connects defects to decision risk mechanisms, reflecting the paper’s positioning that quality failure is not merely technical.

Prevent

Prevention is the first phase because defect removal downstream is typically more costly than defect avoidance upstream, particularly where replication has already occurred. Prevention in constrained environments should be understood as process and representation design rather than as sophisticated automation. Examples include ensuring that key identifiers are required at intake, enforcing consistent date formats, limiting free text where controlled vocabularies are feasible, and implementing basic validation rules. Prevention also includes reducing unnecessary data capture. When forms are overly complex, staff may respond by entering placeholders, skipping fields, or deferring updates, which increases incompleteness and delay.

Consider a primary health facility that records patient encounters. If patient identity is not stabilised through a consistent identifier and basic deduplication checks at registration, the system will accumulate duplicated records that later distort counts of service uptake and

treatment follow-up. Similarly, in SME finance, if loan applications allow multiple spellings and identifiers for the same customer, portfolio risk reporting can be distorted, and collection decisions can become inconsistent. These are not rare edge cases. They are predictable outcomes of capture designs that do not match the realities of frontline work. Prevention in PRO DQ therefore emphasises simple controls that reduce the likelihood of defect creation under throughput pressure.

Responsibilise

Responsibilise addresses a common governance failure: the absence of clear accountability for the quality of specific datasets and definitions. Governance frameworks emphasise stewardship and decision rights because quality work cannot be sustained if responsibility is diffuse (DAMA International, 2017; Khatri & Brown, 2010). In resource-constrained settings, the challenge is to define responsibility without assuming large governance offices. Responsibility therefore focuses on naming owners for critical data domains, defining steward responsibilities for monitoring and definition management, and establishing escalation routes for resolving conflicts.

Responsibility is especially important for semantic defects, such as inconsistent definitions of “active beneficiary,” “completed repair,” or “delivered parcel.” When definitions drift, teams optimise different targets, and dashboards lose credibility. Assigning responsibility for definitions, and for approving changes to definitions, can reduce drift. The governance logic here aligns with “fitness for use” theory because the responsible actor must understand what the data is used for and which decision depends on it (Wang & Strong, 1996). Responsibility also improves auditability, which is vital in public administration and social housing where decisions are subject to scrutiny.

Observe

Observation refers to regular, routine visibility into defect patterns. The goal is not exhaustive measurement but decision-relevant measurement. In constrained environments, observation can be lightweight but must be systematic. It can involve periodic data profiling, exception reports, and a small set of agreed indicators for each critical dataset. Observation is required because many defects are normalised. Staff adapt to them through manual fixes, and managers may not see the hidden cost. By making defects visible, observation supports prioritisation, resource allocation for fixes, and sustained learning.

In a logistics context, observation may focus on address completeness, delivery status timeliness, and duplication in consignment IDs, because these defects directly affect dispatch planning and customer communication. In public administration, observation may focus on missing identifiers, inconsistent demographic fields, and backlog delays, because these influence eligibility determinations and service queue management. Observation should also include metadata checks, such as whether code lists are being used consistently and whether “free text” categories are proliferating.

Diagnose

Diagnosis connects observed defects to root causes. Diagnosis is the phase where ICT4D insights are most salient because defects often emerge from design actuality gaps and local improvisation (Heeks, 2002). Diagnosis should consider not only technical causes but also workload pressure, incentive structures, training, staff turnover, and informal practices. For

example, recurring incompleteness in a housing repairs system may reflect that technicians do not have time to update the system on site and later rely on a spreadsheet compiled by an administrator. The defect is therefore a process design issue rather than a user compliance issue. Diagnosis should also examine whether a defect is created at capture, introduced during transfer, or amplified during reporting and spreadsheet replication.

Diagnosis benefits from tracing the propagation chain. If a dashboard metric is inconsistent across units, diagnosis should trace back through transformations, extract processes, and any parallel reporting sheets. The objective is to identify the smallest set of interventions that prevents recurrence. Diagnosis is also where representational issues arise. Ambiguity often reflects that the representation does not fit the domain, such as a status code set that cannot represent common real-world states. In ontological terms, the system fails to represent reality adequately (Wand & Wang, 1996). Diagnosis should therefore allow changes to the representation, not only cleansing.

Fix

Fix includes both corrective cleaning and systemic correction. Corrective cleaning may be necessary to restore operational continuity, such as deduplicating records for a high priority service list or filling missing values where authoritative sources exist. However, PRO DQ emphasises that cleaning without systemic correction is insufficient because it does not reduce defect generation. Fix therefore includes upstream changes such as improved validation at capture, revised code lists, improved integration routines, and clearer governance rules for edits and approvals.

Fix should also distinguish between immediate fixes and structural fixes. An immediate fix may involve reconciling two lists for a monthly report. A structural fix may involve replacing the need for parallel spreadsheets by improving access controls, adding a simple export that is consistent and versioned, or clarifying that one system is the authoritative source. In SME finance, a fix may include implementing a basic master record process for clients, even if it is manual, with a steward who approves merges and maintains a simple audit log. In health, a fix may involve training combined with form redesign and clearer responsibility for identity resolution.

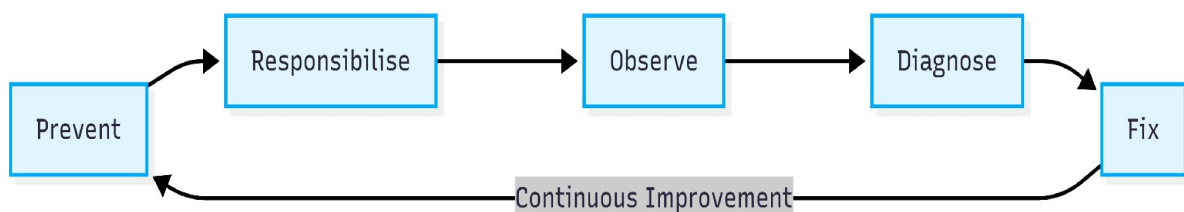


Figure 2: PRO DQ 5 Phase Circular Lifecycle

The circular structure reflects the empirical reality that organisations learn through repeated cycles of observation and correction. It also aligns with governance as a sustained practice rather than a one-time project.

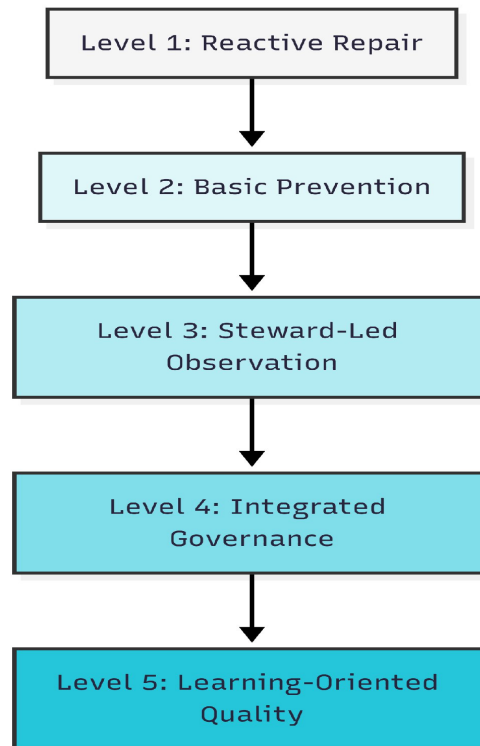


Figure 3: Data Maturity Model for Resource Constrained Environments

Figure 3 provides a realistic maturity path that does not assume full automation. It emphasises routines, stewardship, and learning.

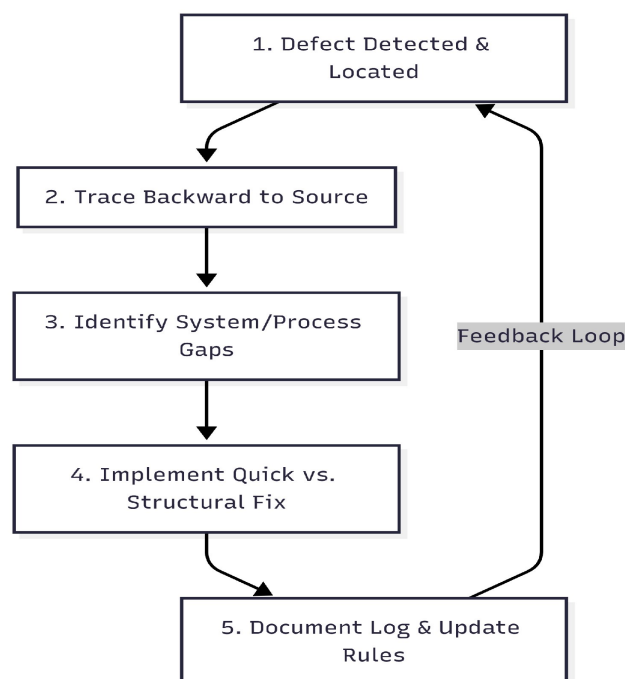


Figure 4: Recursive Root Cause Analysis Workflow

Figure 4 operationalises diagnosis and fix as governance learning. It also treats workarounds as a legitimate object of analysis rather than as noncompliance.

Table 2: Decision Reliability Matrix Linking Defects to Operational Failure

Defect Category	Typical Operational Manifestation	Likely Decision Failure Mode	Sector (Illustrative)	Example
Duplication	two records for same client or asset	double allocation, incorrect prioritisation	health follow-up list	duplicates patients
Incompleteness	missing identifiers or statuses	inability to assess eligibility or risk	housing arrears decisions	lack status dates
Ambiguity	unclear codes and categories	misclassification, inconsistent service	public administration case categories	vary by office
Inconsistency	conflicting values across systems	dashboards disputed, decisions delayed	logistics delivery status	differs across apps
Delay	late updates	missed intervention window, reactive management	SME finance collections	start too late
Workaround drift	parallel spreadsheets become authority	auditability loss, untraceable changes	social program reporting	depends on emailed files

Table 2 highlights that defect categories map to decision failures through specific mechanisms such as misclassification, timing failure, and loss of trust that leads to informal decision making.

Table 3: PRO DQ Framework Implementation Checklist

Phase	Minimum Viable Practices In Constrained Settings	Governance Evidence For Auditability
Prevent	Standardised capture forms; basic validation; unique identifiers for critical entities; controlled code lists for high impact fields	documented standards; versioned forms; validation rules list
Responsibilise	named data owner for each domain; steward for definitions and issue triage; escalation route for conflicts	RACI style role record; definition approval log
Observe	monthly profiling for critical defects; exception reporting; simple defect dashboard tied to decisions	monitoring report; defect backlog with timestamps
Diagnose	trace defects to workflow steps; map workarounds; evaluate design actuality gaps	root cause notes; process maps; change requests
Fix	corrective cleaning for urgent decision products; structural fixes to reduce recurrence; training tied to revised standards	change log; before after metrics; updated controls

Table 3 is intentionally minimal. It supports implementation in organisations that cannot immediately invest in enterprise tooling but can implement routines and documented practices.

Table 4: Quantitative KPIs for Data Quality Success (Technical and Business)

KPI Type	Indicator	Measurement Logic	Decision Relevance
Technical	completeness of critical fields	percent non-missing for predefined elements	reduces manual rework and improves eligibility decisions
Technical	duplication rate	duplicates per entity population using matching rules	prevents double servicing and inflated counts
Technical	consistency across systems	reconciliation pass rate for key fields	reduces dashboard disputes and delays
Technical	timeliness (freshness)	time since last update or event capture latency	improves interventions within decision windows
Business	decision cycle time	time from trigger event to action	indicates whether data supports timely action
Business	rework burden	number of cases requiring manual verification	reflects operational cost of defects
Business	dispute and exception rate	rate of escalations due to data conflicts	indicates trust and governance stability
Business	adoption and trust proxy	proportion of decisions using official dashboards	indicates whether information products are credible

Table 4 separates technical and business indicators to align with the “decision risk pathway” positioning. A purely technical improvement that does not improve decision reliability may not justify the organisational effort.

DISCUSSION AND CONCLUSION

Theoretical Contribution

The manuscript makes a theory-informed contribution by integrating three literatures that are often applied separately. First, it preserves the consumer-oriented foundation that quality is fitness for use (Wang & Strong, 1996) while using the ontological view to clarify why ambiguity and incompleteness are representational failures, not merely administrative inconveniences (Wand & Wang, 1996). Second, it incorporates governance theory through stewardship and decision rights, treating quality as an organisational capability sustained through roles, controls, and monitoring (DAMA International, 2017; Khatri & Brown, 2010). Third, it applies ICT4D theory to explain why standard models underperform when they ignore design actuality gaps and the normality of local improvisation (Heeks, 2002). The core conceptual move is to treat data defects as a decision risk pathway. This move shifts analysis from “how clean is the data” to “how reliable are the decisions that depend on the data, given known defect pathways and governance capacity.”

Managerial Relevance

For managers, PRO DQ offers an implementable sequence that prioritises prevention and accountability before investing in complex tooling. In many constrained environments, organisations adopt dashboards or analytics while the underlying data production routines remain unstable. PRO DQ suggests that analytic credibility depends on stabilising capture, clarifying responsibility for definitions and corrections, and establishing regular visibility into defect patterns. This is not an argument against analytics. It is an argument that analytics trust is an outcome of governance and quality routines.

PRO DQ is also intended to be compatible with mixed maturity. An organisation can begin with one decision-critical dataset, such as patient identity, tenancy status, or client repayment status, and cycle through PRO DQ without waiting for full integration. The circular nature of the model supports incremental learning and avoids the failure mode in which a one-off “cleanup project” is followed by relapse because root causes were not corrected.

Public Policy Implications

At a policy level, the argument supports cautious claims about the digital economy and data-driven development. Institutional discussions increasingly emphasise the importance of data governance for trust and economic coordination (OECD, 2022; World Bank, 2021). However, the practical ability to assure data quality depends on organisational routines and institutional capacity, not only on policy statements. Public sector agencies may benefit from policy support that funds stewardship roles, standard definition work, and cross-agency reconciliation routines. Regional frameworks such as the African Union Data Policy Framework signal the strategic importance of governance, but implementation remains contingent on resourcing, enforcement, and operational fit (African Union, 2022). In this sense, PRO DQ can be read as a micro-level complement to macro-level governance ambitions.

Limitations and Future Research

The limitations are clear. This paper is conceptual design research and does not present field validation or measured impacts. The taxonomy and examples are grounded in theory and plausible organisational mechanisms, but empirical evaluation is required. Future work should test PRO DQ in multi-sector settings, using case studies and pilot interventions, and should measure both defect metrics and decision reliability indicators such as decision cycle time and rework burden. Future research should also examine how organisational incentives, staffing stability, and regulatory oversight shape the effectiveness of responsibility assignment and monitoring.

CONCLUSION

In emerging market settings characterised by manual capture, fragmented systems, spreadsheet mediation, and locally improvised routines, data quality failures should be understood as decision risk pathways. Defects propagate through information products and managerial interpretation into operational error, delayed interventions, and accountability disputes. PRO DQ provides a theoretically grounded and implementable framework that aligns prevention, stewardship, monitoring, diagnosis, and correction into a circular governance logic. The framework does not assume high automation or abundant specialist capacity. Instead, it

emphasises auditable routines and learning-oriented correction that can improve decision reliability under resource constraints.

REFERENCES

- African Union. (2022). *African Union data policy framework*. African Union Commission.
- Batini, C., & Scannapieco, M. (2006). *Data quality: Concepts, methodologies and techniques*. Springer.
- DAMA International. (2017). *DAMA-DMBOK: Data management body of knowledge* (2nd ed.). Technics Publications.
- Heeks, R. (2002). Information systems and developing countries: Failure, success, and local improvisations. *The Information Society*, 18(2), 101–112. <https://doi.org/10.1080/01972240290075039>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152. <https://doi.org/10.1145/1629175.1629210>
- OECD. (2022). *Going digital to advance data governance for growth and well-being*. OECD Publishing.
- Otto, B. (2011). Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, 29, 45–66.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, 45(4), 211–218. <https://doi.org/10.1145/505248.506010>
- Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Engineering Bulletin*, 23(4), 3–13.
- Redman, T. C. (1998). The impact of poor data quality on the typical enterprise. *Communications of the ACM*, 41(2), 79–82. <https://doi.org/10.1145/269012.269022>
- Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. *Communications of the ACM*, 40(5), 103–110. <https://doi.org/10.1145/253769.253804>
- Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), 86–95. <https://doi.org/10.1145/240455.240479>
- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- World Bank. (2021). *World Development Report 2021: Data for better lives*. World Bank. <https://doi.org/10.1596/978-1-4648-1600-0>
- ISO/IEC. (2008). *ISO/IEC 25012:2008 Software engineering: Software product quality requirements and evaluation (SQuaRE): Data quality model*. International Organization for Standardization.
- English, L. P. (2009). *Information quality applied: Best practices for improving business information, processes and systems*. Wiley.
- Loshin, D. (2010). *The practitioner's guide to data quality improvement*. Morgan Kaufmann.