



**Adaa Foundation**  
مؤسسة أداء للدراسات والبحوث

# INSTITUTIONAL PERFORMANCE ASSESSMENT FRAMEWORK FOR WATER UTILITIES

A Diagnostic and Benchmarking Model  
for Evidence-Based Utility Reform in Yemen



## Comprehensive Assessment

Five key pillars covering financial, operational, governance, and data capabilities



## Evidence-Based Benchmarking

Data-driven scoring and comparison for informed decision-making



## Diagnostic & Actionable

Identifies root causes and prioritizes high-impact interventions



## Adaptable & Practical

Designed for utilities in fragile and conflict-affected environments



## Pathway to Reform

Clear roadmap to improve performance and service sustainability



📍 Aden, Yemen

📅 December-2025

**Adaa Foundation for Studies and Research**

# **Institutional Performance Assessment Framework for Water Utilities in Yemen**

*A Diagnostic and Benchmarking Model for Evidence-Based Utility Reform in Yemen*

Prepared by: Hamzah AlGunaid

Aden, Yemen | 2025

*This edition integrates the Institutional Maturity Rubric, revised Donor Instrument Mapping, and the Output/Outcome Accountability Boundary validated through the LWSCA Revenue Performance Diagnostic (2017–2025) and the LWSCA Data-Driven Performance Management Proposal.*

## Contents

|                                                                 |    |
|-----------------------------------------------------------------|----|
| Executive Summary.....                                          | 3  |
| 1. Purpose and Rationale .....                                  | 4  |
| 2. Foundational Principles .....                                | 4  |
| 2.1 Disaggregation Before Diagnosis .....                       | 4  |
| 2.2 Structural vs. Operational Root Causes .....                | 4  |
| 2.3 Statistical Validation Over Narrative Assertion .....       | 4  |
| 2.4 Sequenced, Not Simultaneous, Intervention.....              | 5  |
| 2.5 Output/Outcome Accountability.....                          | 5  |
| 3. Framework Architecture: The Five Pillars.....                | 5  |
| 3.1 Pillar 1: Revenue & Financial Sustainability.....           | 5  |
| Key Indicators.....                                             | 5  |
| Benchmark Thresholds (System-Wide CE) .....                     | 6  |
| Institutional Maturity Rubric — Pillar 1 .....                  | 6  |
| 3.2 Pillar 2: Customer Segmentation & Payment Behavior.....     | 7  |
| Key Indicators.....                                             | 7  |
| Diagnostic Flags .....                                          | 7  |
| Institutional Maturity Rubric — Pillar 2 .....                  | 7  |
| 3.3 Pillar 3: Operational Enforcement & Field Capacity .....    | 8  |
| Key Indicators.....                                             | 8  |
| Diagnostic Flags .....                                          | 8  |
| Institutional Maturity Rubric — Pillar 3 .....                  | 9  |
| 3.4 Pillar 4: Institutional Governance & Accountability .....   | 10 |
| Scored Sub-Indicators.....                                      | 10 |
| 3.5 Pillar 5: Data Infrastructure & Decision Intelligence ..... | 11 |
| Key Indicators.....                                             | 11 |
| Institutional Maturity Rubric — Pillar 5 .....                  | 12 |
| 3.6 The Output/Outcome Accountability Boundary .....            | 14 |
| 4. The Composite Institutional Performance Index (IPI) .....    | 15 |
| Scoring Procedure.....                                          | 15 |
| Maturity Tiers .....                                            | 15 |
| 5. Application Methodology .....                                | 16 |
| 5.1 Five-Phase Sequence .....                                   | 16 |
| 5.2 Data-Adequacy Tiers .....                                   | 16 |
| 5.3 Data Scope and Validation.....                              | 17 |
| 6. Validation Case Study: LWSCA .....                           | 18 |
| 7. Strategic Recommendations Pathway .....                      | 19 |
| 8. Limitations and Considerations .....                         | 20 |
| Appendix A: Full Indicator Reference Table.....                 | 21 |

## **Executive Summary**

---

Water utilities across Yemen and comparable conflict-affected or fragile-state contexts face a recurring diagnostic problem: revenue and service-delivery crises are well documented, but rarely disaggregated into the parts that are structural and therefore require political or institutional resolution, and the parts that are operational and therefore fixable through management action within months. Without this distinction, reform efforts default to uniform, system-wide interventions that misallocate scarce capacity.

This framework generalizes the diagnostic methodology developed through a comprehensive 102-month performance diagnostic of a public water utility in Aden into a reusable, sector-wide assessment model. It is built around five institutional pillars, a composite scoring index, a five-phase diagnostic methodology, and a fully specified institutional maturity rubric that makes the composite index independently replicable by any assessor on any utility.

This edition also introduces a tiered data-adequacy pathway, from a minimum viable three-year snapshot up to the 60+ month depth associated with best-practice diagnostic rigor so that utilities with more limited data histories can still be assessed, at appropriately calibrated confidence; a revised donor-instrument mapping showing which financing mechanism suits each priority; and an explicit Output/Outcome Accountability Boundary that separates what a technical assistance provider delivers from what a utility itself must sustain.

The framework is not a replacement for utility-specific diagnostics. It is a standard lens through which such diagnostics should be structured, so that findings are comparable across utilities, defensible to donors, and actionable to management.

## 1. Purpose and Rationale

---

Public water utilities in fragile and conflict-affected settings are frequently assessed using a single aggregate metric — most often system-wide Collection Efficiency (CE) that obscures more than it reveals. A single CE figure cannot distinguish a utility failing because its customers cannot afford tariffs from one failing because its own suspension policy is administratively eliminating paying customers, nor from one failing because a government ministry simply does not pay its bills.

This framework exists to correct that diagnostic gap. It is built on three propositions, each validated through direct field diagnostic work on a Yemeni public utility:

- Aggregate performance metrics conceal sector-specific dynamics that require fundamentally different interventions.
- A meaningful share of institutional underperformance is operational and reversible that not a fixed consequence of conflict or economic conditions.
- Rigorous, replicable diagnostic methodology — not narrative assessment — is what makes findings credible to donors, ministries, and utility boards alike.

*A utility's revenue crisis is rarely one problem. It is usually three or four distinct problems wearing the same collection-efficiency number.*

Because most public water utilities in Yemen share comparable operating conditions, currency devaluation, fragmented governance, degraded field-enforcement capacity, and structurally similar billing-system architecture, a framework calibrated against one rigorously documented utility is a defensible starting point for sector-wide application, not an overfit single-case exercise. Sections 5 and 6 set out how the framework accommodates utilities with less data than the founding case.

## 2. Foundational Principles

---

### 2.1 Disaggregation Before Diagnosis

No utility performance figure should be interpreted at the system-wide level until it has first been broken down by customer segment (typically domestic, commercial, and governmental/institutional accounts). System-wide averages routinely mask segments performing near international benchmarks alongside segments in acute crisis.

### 2.2 Structural vs. Operational Root Causes

Every underperformance driver identified must be classified as either structural (currency devaluation, conflict-driven institutional fragmentation, governmental non-payment requiring political resolution) or operational (policy design flaws, enforcement capacity gaps, data and reporting weaknesses). Only operational drivers should anchor utility-level intervention plans; structural drivers should be escalated to the appropriate governance level.

### 2.3 Statistical Validation Over Narrative Assertion

Where feasible, relationships between operational levers (e.g., field enforcement intensity) and outcomes (e.g., collection efficiency) should be tested statistically correlation, structural break, and cointegration analysis are the minimum standard rather than assumed from institutional folklore or anecdote.

## 2.4 Sequenced, Not Simultaneous, Intervention

Recommendations should be sequenced by cost, implementation feasibility, and evidence strength beginning with zero-cost policy corrections and progressing toward capital-intensive infrastructure investment rather than presented as an undifferentiated list.

## 2.5 Output/Outcome Accountability

Where a utility's assessment coincides with an active or recently completed donor-funded engagement, the framework distinguishes between outputs delivered by an external partner and outcomes sustained by the utility itself. This distinction, detailed in Section 3.6, ensures institutional scores reflect the utility's own performance rather than a partner's delivery.

# 3. Framework Architecture: The Five Pillars

The framework assesses institutional performance across five pillars. Each pillar is scored independently on a 0–100 scale — using the Institutional Maturity Rubric embedded within each pillar below — and combined into a single Institutional Performance Index (IPI), detailed in Section 4.

| Pillar                                         | Core Question                                                                                                     | Weight in IPI |
|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|---------------|
| 1. Revenue & Financial Sustainability          | Is the utility converting billed services into collected revenue, and is its debt trajectory sustainable?         | 30%           |
| 2. Customer Segmentation & Payment Behavior    | Do distinct customer segments require distinct strategies, and is that distinction currently reflected in policy? | 15%           |
| 3. Operational Enforcement & Field Capacity    | Does non-payment carry a credible operational consequence, and is enforcement capacity adequate?                  | 20%           |
| 4. Institutional Governance & Accountability   | Do internal policies, SOPs, and inter-agency arrangements support or undermine revenue performance?               | 20%           |
| 5. Data Infrastructure & Decision Intelligence | Can the utility generate, trust, and act on its own operational data?                                             | 15%           |

*How to use the rubric: each pillar below lists its key indicators and diagnostic flags (unchanged from the original framework), followed by a 1–5 institutional maturity ladder for each sub-indicator. Score each sub-indicator 1–5 against the described institutional state, average the sub-indicator scores, and multiply by 20 to produce the pillar's 0–100 score used in the IPI formula (Section 4). Worked example: three sub-indicators scored 2, 3, and 2 average to 2.33 → pillar score =  $2.33 \times 20 = 46.7 / 100$ .*

## 3.1 Pillar 1: Revenue & Financial Sustainability

This pillar measures the utility's core financial performance: how much of what it bills it actually collects, and whether its debt position is improving or deteriorating over time.

### Key Indicators

| Indicator                  | Formula                                                        | Data Source                |
|----------------------------|----------------------------------------------------------------|----------------------------|
| Collection Efficiency (CE) | $(\text{Paid Amount} \div \text{Total Billed}) \times 100$     | Billing system, monthly    |
| Revenue Efficiency (RE)    | $\text{Total Paid} \div \text{Total Consumption (m}^3\text{)}$ | Billing + metering records |

|                              |                                               |                        |
|------------------------------|-----------------------------------------------|------------------------|
| Debt Growth Rate             | Year-over-year % change in outstanding debt   | Billing system, annual |
| Revenue Gap to Benchmark     | Total Billed $\times$ (85% – Current CE)      | Calculated             |
| Debt-to-Annual-Billing Ratio | Outstanding Debt $\div$ Total Annual Billings | Calculated             |

### Benchmark Thresholds (System-Wide CE)

| Tier                        | CE Range  | Interpretation                                                             |
|-----------------------------|-----------|----------------------------------------------------------------------------|
| Critical                    | Below 40% | Acute revenue crisis; immediate diagnostic and intervention required       |
| Fragile                     | 40% – 60% | Significant, addressable performance gaps                                  |
| Developing                  | 60% – 70% | Approaching minimum viable operation; consolidation of recent gains needed |
| Sustainability Threshold    | 70% – 75% | Minimum viable operating threshold in constrained environments             |
| International Best Practice | 85% – 95% | Well-functioning utility benchmark                                         |

## Institutional Maturity Rubric — Pillar 1

### 1.1 Collection Efficiency Level & Trend

| Score | Institutional State                                                                                                |
|-------|--------------------------------------------------------------------------------------------------------------------|
| 1     | System CE below 40% and declining year-on-year; no segment disaggregation practiced.                               |
| 2     | System CE below 40%, disaggregated by segment, but trend not systematically tracked against targets.               |
| 3     | System CE 40–60%; segment-specific CE tracked; trend flat or modestly improving.                                   |
| 4     | System CE 60–75%; CE tracked and tied to management accountability; sustained improvement over 12+ months.         |
| 5     | System CE above 75%, sustained, with forward-looking scenario planning informing tariff and enforcement decisions. |

### 1.2 Debt Trajectory & Management

| Score | Institutional State                                                                                                        |
|-------|----------------------------------------------------------------------------------------------------------------------------|
| 1     | Outstanding debt growing faster than billings; no write-off, restructuring, or arrears-relief mechanism exists.            |
| 2     | Debt growth tracked annually; no active management response.                                                               |
| 3     | Debt growth rate monitored by segment; arrears-relief or payment-plan mechanism exists but is ad hoc.                      |
| 4     | Debt growth decelerating; structured arrears-relief and restructuring mechanisms operating with defined eligibility rules. |
| 5     | Debt-to-billing ratio stable or improving; institutional debt-management policy reviewed and updated on a defined cycle.   |

### 1.3 Revenue Gap & Benchmark Awareness

| Score | Institutional State                                                       |
|-------|---------------------------------------------------------------------------|
| 1     | Utility does not calculate or reference the revenue gap to any benchmark. |
| 2     | Revenue gap calculated once, not maintained or updated.                   |

|   |                                                                                                                   |
|---|-------------------------------------------------------------------------------------------------------------------|
| 3 | Revenue gap to sustainability benchmark (70–75% CE) calculated periodically and reported internally.              |
| 4 | Revenue gap tracked routinely and used to justify budget or tariff decisions.                                     |
| 5 | Revenue gap embedded in forward planning, tied to a costed, sequenced recovery pathway with monitored milestones. |

## 3.2 Pillar 2: Customer Segmentation & Payment Behavior

This pillar tests whether the utility's customer base behaves as a single population or as distinct sub-populations requiring differentiated strategy — and whether current policy reflects that distinction.

### Key Indicators

| Indicator                      | Formula / Description                                                                                   | Data Source          |
|--------------------------------|---------------------------------------------------------------------------------------------------------|----------------------|
| Segment-Level CE               | CE calculated separately for domestic, commercial, governmental accounts                                | Billing system       |
| Payment Commitment Ratio (PCR) | % of customer base making $\geq 1$ payment in the period                                                | Billing system       |
| Payment Depth                  | Payment amount $\div$ average monthly bill (months of bill cleared per payment)                         | Billing system       |
| Debt Concentration Ratio       | % of total debt held by a segment $\div$ % of accounts in that segment                                  | Calculated           |
| Structural Break Presence      | Statistical test (e.g., Chow test) for shifts in payment behavior around policy events (tariff changes) | Time series analysis |

### Diagnostic Flags

- A gap of 30+ percentage points in CE between the best- and worst-performing segments indicates uniform system-wide interventions will fail.
- A debt concentration ratio above 10 $\times$  (e.g., a segment holding 50% of debt from under 5% of accounts) signals an institutional, not capacity-based, payment failure.
- A statistically confirmed structural break coinciding with a tariff or policy change indicates the policy itself, not customer behavior in the abstract, is the root cause.

## Institutional Maturity Rubric — Pillar 2

### 2.1 Segment-Level CE Disaggregation Practice

| Score | Institutional State                                                                                                                      |
|-------|------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Performance reported only as a single system-wide CE figure.                                                                             |
| 2     | Segment-level CE calculated but not routinely reported or reviewed.                                                                      |
| 3     | Segment-level CE (domestic/commercial/governmental) reported regularly; no segment-specific targets set.                                 |
| 4     | Segment-specific targets and strategies exist and are reviewed at least quarterly.                                                       |
| 5     | Segment strategy is embedded in tariff, enforcement, and re-engagement design, with documented differentiated interventions per segment. |

## 2.2 Payment Commitment Ratio (PCR) Monitoring

| Score | Institutional State                                                                     |
|-------|-----------------------------------------------------------------------------------------|
| 1     | PCR not calculated.                                                                     |
| 2     | PCR calculated once as part of a one-off diagnostic.                                    |
| 3     | PCR tracked periodically by segment, not linked to intervention design.                 |
| 4     | PCR trends inform customer re-engagement or arrears-relief program design.              |
| 5     | PCR is a standing KPI reviewed by management with segment-specific improvement targets. |

## 2.3 Structural Break / Policy-Impact Testing

| Score | Institutional State                                                                                                            |
|-------|--------------------------------------------------------------------------------------------------------------------------------|
| 1     | No testing of whether tariff or policy changes affected payment behavior.                                                      |
| 2     | Anecdotal or narrative assessment only of policy impact on payment behavior.                                                   |
| 3     | Basic before/after comparison conducted for major tariff changes.                                                              |
| 4     | Statistical testing (e.g., Chow test) applied to confirm structural breaks around policy events.                               |
| 5     | Structural-break testing is a standard step before and after any tariff or policy change, feeding directly into policy design. |

## 3.3 Pillar 3: Operational Enforcement & Field Capacity

This pillar assesses whether the utility has the operational capacity to make non-payment carry a credible consequence, and whether that capacity is being deployed effectively.

### Key Indicators

| Indicator                              | Formula / Description                                                            | Data Source          |
|----------------------------------------|----------------------------------------------------------------------------------|----------------------|
| Enforcement Intensity                  | Disconnections per 1,000 valid accounts, monthly                                 | Field operations log |
| Customer Response Rate                 | % of disconnected customers who subsequently pay and reconnect                   | Field operations log |
| Suspended Account Ratio                | Suspended accounts ÷ total accounts, by segment                                  | Billing system       |
| Illegal-Connection Inspection Coverage | Accounts inspected ÷ total accounts, annualized                                  | Field operations log |
| Enforcement–CE Correlation             | Pearson correlation coefficient between enforcement intensity and CE, by segment | Statistical analysis |

### Diagnostic Flags

- A customer response rate above 70% confirms enforcement is an effective lever where deployed, the constraint is capacity, not policy design.
- A high suspended-account ratio (>20%) in a segment that otherwise shows strong active-account CE indicates policy dysfunction (e.g., rigid auto-suspension), not payment incapacity.

- A statistically insignificant or negative enforcement–CE correlation in a given segment (typically governmental) indicates enforcement is not the appropriate lever for that segment.

### **Institutional Maturity Rubric — Pillar 3**

#### **3.1 Enforcement Intensity vs. Historical Peak**

| Score | Institutional State                                                                                                     |
|-------|-------------------------------------------------------------------------------------------------------------------------|
| 1     | Enforcement intensity below 30% of historical peak or peak intensity unknown.                                           |
| 2     | Enforcement intensity 30–50% of historical peak; no restoration plan in place.                                          |
| 3     | Enforcement intensity 50–75% of historical peak; restoration plan exists but underfunded.                               |
| 4     | Enforcement intensity 75–100% of historical peak; restoration resourced and on schedule.                                |
| 5     | Enforcement intensity at or above historical peak, deployed using response-rate targeting rather than uniform coverage. |

#### **3.2 Customer Response Rate to Enforcement**

| Score | Institutional State                                                                                              |
|-------|------------------------------------------------------------------------------------------------------------------|
| 1     | Response rate not tracked, or below 40%.                                                                         |
| 2     | Response rate tracked; 40–60%.                                                                                   |
| 3     | Response rate 60–70%; used descriptively but not for targeting decisions.                                        |
| 4     | Response rate above 70%, confirming enforcement as an effective lever; used to prioritize deployment by segment. |
| 5     | Response rate above 70% and actively used to calibrate enforcement targeting and forecast recovery impact.       |

#### **3.3 Suspended-Account Ratio Management**

| Score | Institutional State                                                                                                                    |
|-------|----------------------------------------------------------------------------------------------------------------------------------------|
| 1     | Suspended-account ratio above 20% in any segment; suspension is automated with no case-by-case review.                                 |
| 2     | High suspension ratio identified but not attributed to policy design vs. payment incapacity.                                           |
| 3     | Suspension driver diagnosed (policy dysfunction vs. incapacity); no corrective action yet taken.                                       |
| 4     | Auto-suspension triggers removed or under active revision; manual review process being implemented.                                    |
| 5     | Case-by-case suspension review fully operational; suspended-account backlog actively being cleared against a tracked reduction target. |

### 3.4 Pillar 4: Institutional Governance & Accountability

This pillar is assessed through structured document review and management interview rather than transactional data, since governance quality is not directly observable in billing records.

#### Scored Sub-Indicators

| Sub-Indicator                   | Assessment Question                                                                                             |
|---------------------------------|-----------------------------------------------------------------------------------------------------------------|
| SOP Currency & Completeness     | Are billing, suspension, and reconnection procedures documented, current, and consistently applied?             |
| Suspension Policy Design        | Does suspension require case-by-case review, or does an automated trigger remove accounts without verification? |
| Tariff-Setting Responsiveness   | Are tariff changes informed by affordability data, or set independently of payment-capacity evidence?           |
| Inter-Agency Payment Mechanisms | Do formal mechanisms exist to resolve governmental/institutional non-payment beyond utility-level enforcement?  |
| Oversight & Reporting Cadence   | Does a board, ministry, or equivalent body receive and act on regular performance reporting?                    |

#### Institutional Maturity Rubric — Pillar 4

##### 4.1 SOP Currency & Completeness

| Score | Institutional State                                                                                 |
|-------|-----------------------------------------------------------------------------------------------------|
| 1     | Billing, suspension, and reconnection procedures undocumented or severely outdated.                 |
| 2     | SOPs exist for some functions; inconsistently applied.                                              |
| 3     | SOPs exist for all core billing/collection functions; applied with gaps.                            |
| 4     | SOPs current, complete, and consistently applied; periodic review cycle defined.                    |
| 5     | SOPs current, complete, consistently applied, and revised based on documented performance evidence. |

##### 4.2 Suspension Policy Design

| Score | Institutional State                                                                                                            |
|-------|--------------------------------------------------------------------------------------------------------------------------------|
| 1     | Fully automated suspension trigger; no case-by-case review at any stage.                                                       |
| 2     | Automated trigger with limited, informal override capability.                                                                  |
| 3     | Manual review required for suspension, but process undocumented or inconsistently followed.                                    |
| 4     | Documented manual review process operational with defined turnaround standard.                                                 |
| 5     | Manual review operational with appeals mechanism, defined response-time standard (e.g., under 48 hours), and tracked outcomes. |

##### 4.3 Tariff-Setting Responsiveness

| Score | Institutional State                                                             |
|-------|---------------------------------------------------------------------------------|
| 1     | Tariff changes set with no reference to affordability or payment-capacity data. |
| 2     | Affordability considered informally, not evidenced.                             |
| 3     | Affordability data collected but not systematically used in tariff design.      |

|   |                                                                                                            |
|---|------------------------------------------------------------------------------------------------------------|
| 4 | Tariff proposals reference collection-efficiency and affordability evidence before adoption.               |
| 5 | Tariff-setting process formally incorporates affordability modeling and post-implementation CE monitoring. |

#### 4.4 Inter-Agency Payment Mechanisms (Governmental Segment)

| Score | Institutional State                                                                                              |
|-------|------------------------------------------------------------------------------------------------------------------|
| 1     | No mechanism exists to address governmental non-payment; issue unaddressed at any level.                         |
| 2     | Governmental non-payment acknowledged internally; no engagement with ministry-level stakeholders.                |
| 3     | Utility has raised governmental debt with the parent ministry; no formal mechanism established.                  |
| 4     | A formal escalation or inter-agency engagement process exists but is not yet yielding measurable payment change. |
| 5     | Formal inter-agency payment mechanism operational, with measurable reduction in governmental arrears growth.     |

#### 4.5 Oversight & Reporting Cadence

| Score | Institutional State                                                                                                                                         |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | No board, ministry, or equivalent body receives regular performance reporting.                                                                              |
| 2     | Reporting exists but is irregular or not reviewed by decision-makers.                                                                                       |
| 3     | Regular reporting cadence established (e.g., quarterly) to an oversight body.                                                                               |
| 4     | Reporting reviewed by oversight body with documented follow-up actions.                                                                                     |
| 5     | Reporting cadence, review, and follow-up are institutionalized (e.g., a standing Performance Steering Committee) with tracked decisions and accountability. |

### 3.5 Pillar 5: Data Infrastructure & Decision Intelligence

This pillar assesses whether the utility can generate the evidence base this framework itself depends on — a precondition for any of the other four pillars to be reliably assessed or improved over time.

#### Key Indicators

| Indicator                   | Formula / Description                                                                | Data Source                 |
|-----------------------------|--------------------------------------------------------------------------------------|-----------------------------|
| Data Completeness           | Months with usable, reconciled billing data ÷ total months in review period          | Data audit                  |
| Granularity                 | Individual account-level data available vs. aggregate-only reporting                 | System architecture review  |
| Reporting Cadence           | Frequency at which KPIs are compiled and reviewed by management                      | Interview / document review |
| Dashboard / BI Availability | Presence of a functioning performance dashboard accessible to decision-makers        | System review               |
| Forecasting Capability      | Whether the utility can project performance trajectories under alternative scenarios | Capability assessment       |

## Institutional Maturity Rubric — Pillar 5

### 5.1 Data Completeness

| Score | Institutional State                                                                              |
|-------|--------------------------------------------------------------------------------------------------|
| 1     | Fewer than 12 months of usable, reconciled billing data available.                               |
| 2     | 12–24 months of usable data; significant gaps or reconciliation issues.                          |
| 3     | 24–36 months of usable data with minor gaps.                                                     |
| 4     | 36+ months of usable, reconciled data (Tier A equivalent), enabling full statistical validation. |
| 5     | 36+ months of usable data, continuously reconciled and quality-checked as a standing process.    |

### 5.2 Granularity

| Score | Institutional State                                                                                                       |
|-------|---------------------------------------------------------------------------------------------------------------------------|
| 1     | Aggregate-only reporting; no segment or account-level detail available.                                                   |
| 2     | Aggregate data by segment only; no account-level records.                                                                 |
| 3     | Account-level data exists for some functions (e.g., billing) but not others (e.g., field operations).                     |
| 4     | Account-level data available across billing and field operations, not yet integrated.                                     |
| 5     | Integrated account-level dataset across billing, field operations, and payment history, supporting cohort-level analysis. |

### 5.3 Reporting Cadence

| Score | Institutional State                                                                                                          |
|-------|------------------------------------------------------------------------------------------------------------------------------|
| 1     | KPIs compiled irregularly or only on external request.                                                                       |
| 2     | KPIs compiled annually.                                                                                                      |
| 3     | KPIs compiled quarterly, reviewed informally.                                                                                |
| 4     | KPIs compiled monthly and reviewed by management on a defined schedule.                                                      |
| 5     | KPIs compiled monthly (or more frequently), reviewed by a dedicated performance unit, and used to trigger corrective action. |

### 5.4 Dashboard / BI Availability

| Score | Institutional State                                                                                                                       |
|-------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 1     | No dashboard or BI tool exists; all reporting is manual/ad hoc.                                                                           |
| 2     | Static reports produced manually (e.g., spreadsheets) with no live dashboard.                                                             |
| 3     | Basic dashboard exists but is not regularly updated or widely accessible.                                                                 |
| 4     | Functioning dashboard accessible to decision-makers, updated on a defined cadence.                                                        |
| 5     | Real-time or near-real-time dashboard with seasonal adjustment and drill-down by segment, embedded in routine management decision-making. |



### 5.5 Forecasting Capability

| Score | Institutional State                                                                                                         |
|-------|-----------------------------------------------------------------------------------------------------------------------------|
| 1     | No forecasting capability; performance discussed only in retrospect.                                                        |
| 2     | Simple trend extrapolation used informally.                                                                                 |
| 3     | Basic scenario modeling (e.g., regression-based) produced occasionally.                                                     |
| 4     | Validated time-series forecasting (e.g., ARIMA/SARIMA) used for planning, produced periodically.                            |
| 5     | Forecasting capability institutionalized: refreshed on a defined cycle, used to set targets, and validated against actuals. |

### 3.6 The Output/Outcome Accountability Boundary

Pillar 3 (Operational Enforcement) and Pillar 4 (Institutional Governance) both mix elements that a utility can directly control with elements that depend on external actors, donors, technical partners, or ministry-level counterparts. Scoring these pillars without separating the two produces two failure modes: (a) utility management disputes a low score for a gap they argue is not theirs to close, and (b) a donor-funded project is credited or penalized for outcomes outside its contractual scope.

The framework adopts the following accountability boundary, validated through the LWSCA Data-Driven Performance Management engagement design:

- Outputs are the direct, contracted deliverables of a technical assistance or capital engagement — systems built, dashboards deployed, KPI frameworks designed, staff trained, SOPs drafted. Responsibility for outputs sits with the delivery partner and should be scored as a Pillar 5 (Data Infrastructure) or Pillar 4 (Governance design) maturity gain once delivered and verified.
- Outcomes are the downstream institutional and financial results that depend on the utility's own sustained execution, enforcement actually restored, auto-suspension policy actually revised and applied, arrears-relief programs actually run, ministries actually engaged. Responsibility for outcomes sits with utility management and should be scored as Pillar 1–3 maturity gains only once field evidence confirms the change.
- A completed output (e.g., a dashboard delivered and functioning) should not, by itself, raise a Pillar 1–3 score. It raises Pillar 5 only. The corresponding Pillar 1–3 gain is scored later, and separately, once the utility's own execution produces observable results.
- Where a governance gap depends on an external body outside the utility's authority (e.g., a governmental ministry's own budget disbursement process), Principle 2.2 governs: the gap is noted and escalated, but a utility should not be scored down on Pillar 4 for a mechanism it has no unilateral authority to create — only for failing to initiate or pursue engagement toward one.

This boundary should be applied consistently whenever the framework is used to assess a utility with an active or recently completed donor-funded engagement, to ensure the IPI reflects the utility's own institutional performance rather than the performance of an external delivery partner.

## 4. The Composite Institutional Performance Index (IPI)

Each pillar is scored 0–100 using the maturity rubric in Section 3. The Institutional Performance Index (IPI) is the weighted sum of the five pillar scores:

$$\text{IPI} = (0.30 \times \text{P1}) + (0.15 \times \text{P2}) + (0.20 \times \text{P3}) + (0.20 \times \text{P4}) + (0.15 \times \text{P5})$$

### Scoring Procedure

- Score every sub-indicator listed under each pillar in Section 3, 1–5, against the described institutional state.
- Average the sub-indicator scores within each pillar.
- Multiply the pillar average by 20 to convert to the pillar's 0–100 score.
- Apply the IPI formula above to the five pillar scores to produce the composite index.

*This procedure replaces narrative or judgment-only scoring with a documented, auditable path from raw institutional evidence to the composite index, making the IPI independently reproducible by a second assessor on a second utility.*

Weights reflect the relative diagnostic power observed across utility assessments to date — Revenue & Financial Sustainability and Operational Enforcement / Governance carry the highest weight because they are both the most consequential to fiscal outcomes and the most directly actionable by utility management within a 12–18 month horizon.

### Maturity Tiers

| Tier          | IPI Range | Characteristics                                                                                                  |
|---------------|-----------|------------------------------------------------------------------------------------------------------------------|
| Crisis        | 0 – 30    | System-wide revenue collapse; debt growth outpacing collections; minimal enforcement or governance controls      |
| Fragile       | 31 – 50   | Partial functioning with major structural or policy gaps; recovery requires sequenced, multi-pillar intervention |
| Transitional  | 51 – 65   | Identifiable, evidenced recovery pathway; reforms underway or recently initiated                                 |
| Stable        | 66 – 80   | Approaching regional sustainability thresholds (70–75% CE equivalent); consolidation phase                       |
| Best Practice | 81 – 100  | Meets or exceeds international utility benchmarks across all pillars                                             |

## 5. Application Methodology

Each pillar assessment follows a consistent five-phase analytical sequence, adapted from the methodology validated in the founding utility diagnostic underlying this framework.

### 5.1 Five-Phase Sequence

| Phase                                    | Activity                                                                                         | Primary Output                     |
|------------------------------------------|--------------------------------------------------------------------------------------------------|------------------------------------|
| 1. Data Foundation                       | Construct an enhanced monthly dataset integrating billing, field operations, and policy records  | Structured multi-year dataset      |
| 2. Descriptive Analysis                  | Systematic comparison across time periods, customer segments, and operational dimensions         | Comparative analytical tables      |
| 3. Operational Analysis                  | Assess field enforcement intensity, response rates, and account status distribution              | Enforcement effectiveness profile  |
| 4. Diagnostic Analysis                   | Statistical root-cause testing: correlation, structural break, cointegration analysis by segment | Validated root-cause findings      |
| 5. Predictive Modeling & Recommendations | Scenario-based projections and sequenced, costed intervention roadmap                            | Recovery pathway with ROI estimate |

### 5.2 Data-Adequacy Tiers

The five-phase sequence above assumes data depth comparable to the founding LWSA diagnostic. Most utilities beginning an assessment will not have this depth. The following tiers allow the framework to be applied, at appropriately calibrated confidence, regardless of starting data maturity, so that a utility with limited records is not excluded from assessment but is instead directed toward Pillar 5 remediation as its own first priority. Sixty months (five years) of usable monthly data is treated as the best-practice threshold: below this depth, seasonal and cointegration testing lose statistical reliability; at or above it, the full diagnostic toolkit validated in the founding LWSA study, including structural break, cointegration, Granger causality, and seasonal (SARIMA) forecasting that becomes trustworthy.

| Data Tier                | Minimum Requirement                                                                      | What Can Be Scored                                                                                                                                          | Confidence Level |
|--------------------------|------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| Tier A+ (Best Practice)  | 60+ months, account-level or near-account-level data, field operations log               | All five pillars; full statistical validation including cointegration, Granger causality, and seasonal (SARIMA) forecasting                                 | Very High        |
| Tier A (Robust)          | 36–60 months, near-account-level or well-reconciled aggregate data, field operations log | All five pillars; correlation, structural break (Chow), and non-seasonal (ARIMA) forecasting validated; cointegration/Granger results treated as indicative | High             |
| Tier B (Typical utility) | 24–36 months aggregate billing, partial field records                                    | Pillars 1, 2, 4 fully; Pillar 3 partially (intensity only, no correlation testing); Pillar 5 scored low by default                                          | Moderate         |
| Tier C (Minimal)         | Current-year snapshot only, no time series                                               | Pillar 1 (point-in-time CE); Pillar 4 (interview-based); Pillars 2/3/5                                                                                      | Indicative only  |

|  |  |                                                                   |  |
|--|--|-------------------------------------------------------------------|--|
|  |  | flagged as insufficient data —<br>priority for Phase 1 investment |  |
|--|--|-------------------------------------------------------------------|--|

*A Tier C or Tier B outcome is itself a diagnostic finding: it identifies Pillar 5 (Data Infrastructure) as the utility's own entry-point priority, consistent with Priority P4 in Section 7. LWSCA's own 102-month dataset exceeds even the Tier A+ threshold, illustrating the ceiling of what sustained data investment ultimately enables rather than the floor most utilities should expect to start from."*

**5.3 Data Scope and Validation**

The founding diagnostic utilized utility billing-system monthly aggregate data by usage category. Individual customer-level transaction records were not available in that case; findings derived from statistical pattern analysis of aggregate trends. This aggregation level enables robust time-series analysis and sectoral comparison while precluding individual customer behavior modeling, a constraint that applies with even greater force to Tier B and Tier C utilities and should be documented transparently in any assessment report.

Validation is best achieved through triangulation across billing-system statistics, field-operations records, and international benchmarking comparisons, regardless of data tier.

## 6. Validation Case Study: Local Water & Sanitation Corporation (LWSCA)

The framework is illustrated below through retrospective application to the Local Water and Sanitation Corporation Authority – Aden (LWSCA), the utility whose 102-month diagnostic (2017–2025) informed the framework's design. The pillar scores below are Adaa's interpretive application of the Section 3 rubric to the diagnostic's published findings, presented as illustrative pending a full re-scoring exercise using the finalized 1–5 ladders in Section 3, which is recommended as the next validation step before the framework is applied to a second utility.

| Pillar                                         | Weight | Illustrative Score | Weighted Contribution | Basis                                                                                                                                                       |
|------------------------------------------------|--------|--------------------|-----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Revenue & Financial Sustainability          | 30%    | 25 / 100           | 7.5                   | System CE of 33.9%, well below the 70–75% sustainability threshold; debt up 233% since 2017                                                                 |
| 2. Customer Segmentation & Payment Behavior    | 15%    | 70 / 100           | 10.5                  | Strong evidence of differentiated segment dynamics correctly identified and statistically validated (structural break confirmed, $p=0.014$ )                |
| 3. Operational Enforcement & Field Capacity    | 20%    | 20 / 100           | 4.0                   | Enforcement intensity collapsed 62–75% from historical peaks, despite an 82.4% customer response rate confirming the lever remains effective                |
| 4. Institutional Governance & Accountability   | 20%    | 30 / 100           | 6.0                   | Rigid auto-suspension policy identified as actively eliminating payment-capable customers; no inter-agency governmental payment mechanism in place          |
| 5. Data Infrastructure & Decision Intelligence | 15%    | 50 / 100           | 7.5                   | 102 months of usable aggregate data enabled rigorous diagnosis, but account-level granularity and real-time dashboards were absent prior to this diagnostic |

### Illustrative Institutional Performance Index (IPI): 35.5 → Fragile Tier

This score is consistent with the diagnostic's own central conclusion: LWSCA's crisis is severe but not structurally hopeless. The high Pillar 2 score reflects strong analytical differentiation of customer segments; the low Pillar 3 and Pillar 1 scores reflect the operational collapse driving the crisis; and the mid-range Pillar 5 score reflects that the data existed to perform this diagnosis but had not previously been organized to support ongoing management decision-making itself one of the diagnostic's own recommendations (Priority P4), now the subject of an active technical assistance engagement (Section 7).

## 7. Strategic Recommendations Pathway

Where a utility's IPI places it in the Crisis or Fragile tier, the framework recommends the following generic sequencing logic, adapted to the utility's specific pillar scores and mapped to the donor financing instrument each priority is best suited to.

| Priority | Target Pillar                | Typical Action                                                                                                   | Donor Instrument                                                                                                    | Cost Profile                     | Timeline          |
|----------|------------------------------|------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|----------------------------------|-------------------|
| P0       | Governance (4)               | Eliminate policy-driven revenue leakage (e.g., rigid auto-suspension, undocumented SOPs)                         | Technical assistance grant — advisory only, no capital                                                              | Zero / low cost                  | 0–1 mo            |
| P1       | Enforcement (3)              | Restore field enforcement capacity toward historical intensity, targeting highest-response segments first        | Results-based financing tied to CE improvement milestones                                                           | Low–medium                       | 1–6 mo            |
| P2       | Segmentation (2)             | Segment-specific re-engagement (affordability relief for domestic; suspension resolution for commercial)         | Blended TA + small capital grant (payment infrastructure, outreach)                                                 | Medium                           | 6–12 mo           |
| P3       | Financial Sustainability (1) | Reassess tariff and subsidy structure against affordability evidence from P2                                     | Policy dialogue support, typically multilateral (e.g., World Bank, GIZ)                                             | Policy-level                     | 12–18 mo          |
| P4       | Data Infrastructure (5)      | Institutionalize dashboards, KPI reporting cadence, predictive monitoring; establish a standing Performance Unit | Capital/systems TA grant — worked example: LWSCA Data-Driven Performance Management Proposal, \$134,130 / 14 months | Donor-funded, one-off + handover | Parallel, 1–24 mo |

*ROI-scope note: a P4-only project ROI (e.g., ~432%, ~5.3:1 in the LWSCA proposal) measures return on the data-infrastructure investment alone and is not directly comparable to the full P1–P3 operational recovery ROI (e.g., 2,500–4,260%+ in the LWSCA diagnostic), which is calculated against total investment across all three operational priorities. Donor materials referencing both figures should state the denominator explicitly.*

Structural drivers identified under Pillar 4 that fall outside utility control — most commonly governmental non-payment — should be excluded from utility-level performance targets and escalated as a distinct workstream for ministry- or cabinet-level engagement, consistent with the founding diagnostic's treatment of governmental debt, and with the Output/Outcome Accountability Boundary in Section 3.6.

## **8. Limitations and Considerations**

---

- The framework depends on the availability of usable billing and operations data; utilities assessed at Data Tier C (Section 5.2) should treat the resulting IPI as indicative only and prioritize Pillar 5 remediation before a full-confidence IPI can be computed.
- Pillar 4 (Governance) relies on structured qualitative scoring via the maturity rubric and is therefore more susceptible to assessor judgment than the largely quantitative Pillars 1–3; independent cross-review of governance scores is recommended.
- The framework measures institutional and operational performance; it does not assess physical infrastructure condition (network age, non-revenue water from physical losses, treatment capacity), which should be evaluated through a complementary engineering assessment.
- Weights in the IPI formula (Section 4) are a starting default calibrated against the founding case study; sector regulators or donor consortia adopting this framework across multiple utilities may recalibrate weights based on a broader validation sample.
- The Section 6 case study remains illustrative pending a full re-scoring of LWSCA against the finalized Section 3 rubric — recommended as the immediate next validation step (see cover note).

## Appendix A: Full Indicator Reference Table

| Pillar | Indicator                       | Formula / Basis                                                       |
|--------|---------------------------------|-----------------------------------------------------------------------|
| 1      | Collection Efficiency (CE)      | $(\text{Paid} \div \text{Billed}) \times 100$                         |
| 1      | Revenue Efficiency (RE)         | $\text{Paid} \div \text{Consumption (m}^3\text{)}$                    |
| 1      | Debt Growth Rate                | YoY % change, outstanding debt                                        |
| 1      | Revenue Gap to Benchmark        | $\text{Billed} \times (85\% - \text{CE})$                             |
| 1      | Debt-to-Billing Ratio           | $\text{Outstanding Debt} \div \text{Annual Billings}$                 |
| 2      | Segment-Level CE                | CE by customer category                                               |
| 2      | Payment Commitment Ratio        | % accounts with $\geq 1$ payment/period                               |
| 2      | Payment Depth                   | $\text{Payment} \div \text{average monthly bill}$                     |
| 2      | Debt Concentration Ratio        | $\% \text{ debt share} \div \% \text{ account share}$                 |
| 2      | Structural Break Presence       | Chow test around policy events                                        |
| 3      | Enforcement Intensity           | $\text{Disconnections} / 1,000 \text{ valid accounts} / \text{month}$ |
| 3      | Customer Response Rate          | % disconnected who pay & reconnect                                    |
| 3      | Suspended Account Ratio         | $\text{Suspended} \div \text{total accounts, by segment}$             |
| 3      | Inspection Coverage             | $\text{Accounts inspected} \div \text{total accounts (annualized)}$   |
| 3      | Enforcement–CE Correlation      | Pearson r, enforcement intensity vs. CE                               |
| 4      | SOP Currency & Completeness     | 1–5 scored review                                                     |
| 4      | Suspension Policy Design        | 1–5 scored review                                                     |
| 4      | Tariff-Setting Responsiveness   | 1–5 scored review                                                     |
| 4      | Inter-Agency Payment Mechanisms | 1–5 scored review                                                     |
| 4      | Oversight & Reporting Cadence   | 1–5 scored review                                                     |
| 5      | Data Completeness               | $\text{Usable months} \div \text{total months}$                       |
| 5      | Granularity                     | Account-level vs. aggregate-only                                      |
| 5      | Reporting Cadence               | KPI review frequency                                                  |
| 5      | Dashboard / BI Availability     | Binary / maturity scale                                               |
| 5      | Forecasting Capability          | Capability assessment                                                 |